

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 11:01:44 ; Search time 16.41 Seconds
(without alignments)
632.398 Million cell updates/sec

Title: US-09-673-395a-238

Perfect score: 564
Sequence: 1 CHARLNTDSSRLAKMLLAVL.....NGLAMHTYDSTWCMKSN 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	195.5	34.7	95	1	prostatic steroid-
2	195.5	13.6	280	1	hypothetical prote
3	76.5	13.5	212	2	D-alanyl-D-alanine
4	73.5	13.0	590	2	unknown protein, 3
5	71.5	12.7	98	1	prostatic steroid-
6	71.5	12.7	131	2	hypothetical prote
7	71.5	12.7	4848	2	pristinamycin 1 sy
8	71.5	12.6	303	2	cell division prot
9	70	12.4	303	2	cell division prot
10	69.5	12.4	457	2	hypothetical prote
11	69.5	12.3	412	2	hypothetical prote
12	69.5	12.3	1048	2	protein FIN19.15 (
13	69.5	12.2	219	2	hypothetical prote
14	69.5	12.2	638	2	hypothetical prote
15	68.5	12.1	173	2	hypothetical prote
16	68.5	12.1	213	2	hypothetical prote
17	68.5	12.1	496	2	hypothetical prote
18	68.5	12.1	561147	2	TCM10 protein - ye
19	68	12.1	516	2	tryptophan halogen
20	68	12.1	1071	2	ATP-binding protei
21	68	12.1	4717	2	indole-3-glycerol-
22	67.5	12.0	268	1	probable membrane
23	67.5	12.0	871	2	hypothetical prote
24	67.5	11.9	738	2	FMG22 protein prec
25	66.5	11.8	95	2	hypothetical prote
26	66.5	11.8	234	2	phosphomethylpyrim
27	66.5	11.8	274	2	nitrate ABC transp
28	66.5	11.8	359	2	cyclin 2b - Arabid
29	66.5	11.8	429	2	

30	66.5	11.8	663	2	T37772
31	66.5	11.8	1601	2	T18800
32	65.5	11.6	162	2	I49124
33	65.5	11.6	429	2	D84554
34	65.5	11.6	490	2	T31646
35	65.5	11.6	603	1	W1ML8
36	65	11.5	137	2	F97763
37	65	11.5	269	2	T04394
38	65	11.5	322	2	E89057
39	65	11.5	335	2	T31712
40	65	11.5	349	2	T15082
41	65	11.5	874	1	S07380
42	65	11.5	975	2	A86258
43	65	11.5	1255	2	T06267
44	65	11.5	2145	2	S61041
45	64.5	11.4	113	2	JC4143

ALIGNMENTS

RESULT 1

prostatic steroid-binding protein chain C3 precursor - rat

N:Alternate names: prostaticin (Norway rat)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Dec-1982 #sequence_revision 14-Nov-1983 #extLchange 28-May-1999
C:Accession: A92395; A92433; A91108; A42392; A03250
R:Parker, M.G.; White, R.; Hurst, H.; Needham, M.; Tilly, R.
J. Biol. Chem. 258, 12-15, 1983

A:Title: Prostatic steroid-binding protein. Isolation and characterization of C3 gene
A:Reference number: A92395; MUID:83082848

A:Accession: A92395

A:Residues: 1-95 <PAR>

A:Cross-references: GB:V01263; GB:J00777; NID:956993; PIDN:CA24577.1; PID:956994
R:Vilskoch, D.H.; Perry, S.T.; Lea, O.A.; Stafford, D.W.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 258, 8861-8866, 1983

A:Title: Isolation of two genomic sequences encoding the M-r = 14,000 subunit of rat

A:Reference number: A92433; MUID:83238526

A:Accession: A92433

A:Molecule type: mRNA

A:Residues: 1-78, 'S', 80-95 <VIS>

R:Peeters, B.; Rombauts, W.; Mous, J.; Heyns, W.
Eur. J. Biochem. 115, 115-121, 1981

A:Title: Structural studies on rat prostatic binding protein. The primary structure o

A:Reference number: A91108; MUID:81188769

A:Accession: A91108

A:Molecule type: protein

R:Tan, J.A.; Marschke, K.B.; Ho, K.C.; Perry, S.T.; Wilson, E.M.; French, F.S.
J. Biol. Chem. 257, 4456-4466, 1992

A:Title: Response elements of the androgen-regulated C3 gene.
A:Reference number: A42392; MUID:92165796

A:Accession: A42392

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-16, 'T', 18-95 <TRAN>

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBI:83851, NCBI:83856)

C:Comment: C3 is encoded by two unique genes that differ from each other only in the

C:Comment: Steroid-binding protein, the principal androgen-dependent secretory protei

ng C2 and C3 chains. The chains of each dimer are linked by disulfide bonds. This pro

C:Superfamily: uteroglobin

C:Keywords: glycoprotein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-95/Product: prostatic steroid-binding protein C3 chain #status experimental <MAT

F:35/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 34.7%; Score 195.5; DB 1; Length 95;
Best Local Similarity 35.8%; Pred. No. 8e-13;

```
Matches      34;    Conservative      30;    Mismatches      30;    Indels      1;    Gaps      1;
```

```
OY      14 MKLLMTMLTALALLHCYAD-SGCKLEDDVYEKTINSIDISPEYKELLQEFIDSDAAAEAM 72  
       |||:: :|| :||| ||| ::||| ::| :|| ::| :  
Db      1 MLKLFLPLVTITICGYASGGCSILIDEVIIRGTINSTVLTHDKYLKVPYOIHFEKEAV 60
```

```
OY      73 GKPKOCFLNOSHRTLTKNFGLMHTHYDSTIWCNMKS 107  
       :|||||:: :||| :||| ::||| :|  
Db      61 KOFKCCEFDQTDLTENVGVMMEATFNSSSCOOPS 95
```

```
RESULT      2  
T38449  
hypothetical protein SPAC27E2.02 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomycetes pombe  
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999  
A:Accession: T38449  
R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z21749  
A:Accession: T38449  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A.Residues: 1-280 <MW>  
A.Cross-references: EMBL:Z898978; PIDN:CAB11676.1; GPSDB:GN00066; SPDB:SPAC27E2.02  
A.Experimental source: strain 97zh-; cosmid c27E2  
C.Genetics:  
A:Gene: SPDH:SPAC27E2.02  
A:Map position: 1  
A:Introns: 185/3; 225/3
```

```

Query Match      13.6%; Score 76.5; DB 2; Length 280;
Best Local Similarity 30.3%; Pred.No.3,1; Mismatches 28; Indels 15; Gaps 3;
Matches    23; Conservative   10;

OY       37 LLEDMVETKINSDISPEKKELLOEPIFSD---AAAEAMGKFKEQ-----CFINQSRT 86
           | :|::||:||||:|| ||| |:| ::|||         ||:|
Db        68 LAEDTLLSVATGDVCITFTSMDLKLVELVIDAEGMAARESKSGDESDKETPYMLNKRSHYV 127

OY       87 LKNFGIMMHTVDYSIW 102
           | :|::||:||||:|| ||| |:| ::|||         ||:|
Db       128 AKT-----PGLDGEPMW 138


RESULT     3
H87530
C:D-alanyl-D-alanine dipeptidase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescutus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H87530
R:Nierman,W.C.; Feldbljum,T.V.; Paulsen,I.T.; Nelson,K.E.; Eisen,J.; Heidelberg,J.B.; Laub,M.T.; Deboy,R.T.; Dodson,R.J.; Durkin,A.S.; Gwin,M.L.; Haft,D.H.; Kolonin,J.; Ermolaeva,M.; White,O.; Salzberg,S.L.; Shapiro,L.; Venter,J.C.; Fraser,C.M.Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescutus.
A:Reference number: A87249; NCID:21173698; PMID:11259647
A:Accession: H87530
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <(STO>
A:Cross-references: GB:E005673; NID:g13423786; PIDN:AAR24244.1; GSPPB:GN00148 C:Genetics
C:Gene: CCZ273
C:Superfamily: Enterococcus faecium transposon TnI546 D-alanine-D-alanine dipeptidase

Query Match      13.5%; Score 76; DB 2; Length 212;
Best Local Similarity 29.9%; Pred.No.2,6;
Matches    26; Conservative   13; Mismatches 16; Indels 32; Gaps 5;

OY       31 ADSCGKILEDVVEKT-----INSDISIFEYE--LLOEFIDSAAAAFAWG 73
           | :|::||:||||:|| ||| |:| ::|||         ||:|
```

```

Db      19  AETGKTIASLDVLYTTTTPGICKIDIRYAGANNFNFGPILEYKSAAYLQR-----PAAEALG 73
OY      74  KFKOCFLNOSHRTL--KNFGIMAHITVY 98
           || | | | | | | | | | |
Db      74  RI-----HRAALAKGYGLLIHDAY 92

RESULT  4
A96656
unknown protein, 38394-36551 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96656
C:Orthologs: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alon
             Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
             ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
             C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzita
             Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
             ker, M.; Wu, D.; Yiu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: A96656
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-590 <STO>
A:Cross-references: GB:AE005173; NID:g10092251; PIDN:AMG12665.1; GSPDB:GN00141
C:Genetics:
A:Gene: p16M19.15
A:Map position: 1

```

```

Query Match          13 0%: Score 73.5; DB 2; Length 590;
Best Local Similarity 32.3%; Pred. No. 14;
Matches 21; Conservative 8; Mismatches 23; Indels 13; Gaps 3;

QY      26 LHCVCADSG-----CKLLDMVEKTIINSISIPYEKKLLOEFIDSDAAEA-----MGK 74
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      291 LISLCNNGRNSDASRLSLDMLKENINPDLYF--FNALIDAFVKEGKLVEAEKLYDEMYK 348
       |||
QY      75 FKOCF 79
       |||
Db      349 SKHCF 353

RESULT    5
BOKRT2
Prostatic steroid-binding protein chain C2 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:date: 15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change 24-Sep-1999
C:Accession: A03251; A26671
R:parker, M.; Needham, M.; White, R.
Nature 298, 92-94, 1982
A:title: Prostatic steroid binding protein: gene duplication and steroid binding.
A:Reference number: A93286; MUID:82220075
A:Accession: A03251
A:Molecule type: mRNA
A:Residues: 1-98 <PAR>
A:Cross-references: GB:J00776; NID:g206448; PIDN:AAA51641.; PID:g206450
R:deleay, B.; Dirrux, V.; Decourt, J.L.; Claessens, F.; Peeters, B.; Kombauts, W.
Nucleic Acids Res. 15, 1627-1641, 1987
A:title: Rat prostatic binding protein: the complete sequence of the C2 gene and its
A:Reference number: A26671; MUID:87146484
A:Accession: A26671
A:Molecule type: DNA
A:Residues: 1-23, 'Q', 26-86, 'I', 88-94, 'WVIQTINPGRWFSEIN' <DEL>
A:Cross-references: GB:X05034; NID:g56857; PIDN:CAA28708.1; PID:g56858
C:Comment: Steroid-binding protein, the principal secretory protein in rat prostatic
The chains of each dimer are linked by disulfide bonds.
C:Superfamily: uteroglobin
C:Keywords: heterotetramer; prostate; steroid binding

```


OY 62 -FIDSDAAAEAMGKPKOCFLNOSHRTLNFGILMHTVYDSIW 102
 DB 53 GILAEAGAKKEGYESLRESFSLSEIPEMRD-----KVSYDEW 89

RESULT 14

T47569
 hypothetical protein F24B22.60 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #lexl_change 20-Apr-2000
 C:Accession: T47569
 R:Blocker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quehler, F.; Salanoubat, M.
 submitted to the Protein Sequence Database, January 2000
 A:Reference number: 223016
 A:Accession: T47569
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-638 <HIO>
 A:Cross-references: EMBL:AL132957
 A:Experimental source: cultivar Columbia; BAC clone F24B22
 C:Genetics:
 A:Map position: 3
 A:Introns: 192/3; 217/1; 240/3; 270/2; 314/3; 392/2; 551/2
 A:Note: F24B22.60

Query Match 12.2%; Score 69; DB 2; Length 638;
 Best Local Similarity 21.7%; Pred. No. 43;
 Matches 23; Conservative 22; Mismatches 33; Indels 28; Gaps 3;

OY 4 RLNTDSSRLAMKILMVLMAALLHGYADSG-----CKLEPMVEKTI NSDISIPRYKE 57
 DB 116 RSGNVRGRITGLMLLVASVFLRYIMGVAVVDHARLKEFVVYRTLDDMSMAQ--- 172
 OY 58 LLOEFIDSDAAAEAM-----GKPKOCFLNOSHRT 86
 DB 173 --RVAENQASQPMRYLEKPIPEIWMQKESGNYKOCVTRPKNYT 216

RESULT 15

T01889
 hypothetical protein F8M12.4 - Arabidopsis thaliana
 N:Alternate names: hypothetical protein F25124.80
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #lexl_change 21-Jul-2000
 C:Accession: T01889; T04282
 R:Madson, C.; Graves, T.; Cotton, M.; Modde, T.
 submitted to the EMBL data library, July 1998
 A:Description: The sequence of A. thaliana F8M12.
 A:Reference number: Z14450
 A:Accession: T01889
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-173 <MAD>
 A:Cross-references: EMBL:AF080118; NID:q3513725; PID:q3513732
 A:Experimental source: cultivar Columbia
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, A.
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15261
 A:Accession: T04282
 A:Molecule type: DNA
 A:Residues: 1-173 <BEV>
 A:Cross-references: EMBL:AL049525
 A:Experimental source: cultivar Columbia; BAC clone F25124
 C:Genetics:
 A:Map position: 4
 A:Note: F8M12.4; F25124.80
 C:Superfamily: Arabidopsis thaliana hypothetical protein F8M12.2

Query Match 12.1%; Score 68.5; DB 2; Length 173;
 Best Local Similarity 24.3%; Pred. No. 12;
 Matches 27; Conservative 17; Mismatches 36; Indels 31; Gaps 6;

OY 24 ALLHGYADSGC-----KLEDMV-----EKTINSDISIPRYKEILLQEFIDSDAAAE 70
 DB 2 APLHRYVSNIGCGKRYTVRFCDKTIYTDNGESRTINEELNSGPTKIKIRDPDEDQCYL 61
 OY 71 AMGKPKOCFLNOSHRTLNKNEGL---MMH--TVYD-----SICGM 105
 DB 62 LNYDGYMLMNDNRG--KNKGIDYDLVHWTAVDDINIDIVDEMKIWCEI 110

Search completed: June 20, 2002, 11:03:07
 Job time: 83 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 11:02:44 ; Search time 11.86 Seconds
(Without alignments)
352.589 Million cell updates/sec

Title: US-09-673-395A-238

Perfect score: 564

Sequence: 1 CHARLNTDSSRLAMKRLNLV.....NFGLMHTVDSIMCKMSKN 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	496	87.9	95	1	MGBB_HUMAN
2	281	49.8	93	1	MGBA_HUMAN
3	195.5	34.7	95	1	PSC3_RAT
4	71.5	12.7	131	1	YRS9_MYCTU
5	70	12.4	303	1	FTSY_RICPR
6	69	12.2	112	1	PSC2_RAT
7	68.5	12.1	684	1	TC10_YEAST
8	68	12.1	1071	1	PR16_YEAST
9	67.5	12.0	268	1	TRPC_ACTCA
10	67.5	12.0	871	1	SC10_YEAST
11	66.5	11.8	234	1	RADC_HAEIN
12	66.5	11.8	663	1	TAZ1_SCHPO
13	65.5	11.6	162	1	IL15_MOUSE
14	65.5	11.6	603	1	VEL_HPV08
15	65	11.5	2144	1	G1T1_YEAST
16	64.5	11.4	113	1	MTH_CAIAT
17	64.5	11.4	153	1	Y156_UREPA
18	64.5	11.4	162	1	VE1_HPV21
19	64.5	11.4	603	1	VE1_HPV21
20	64.5	11.4	611	1	DYNL_ARATH
21	64	11.3	473	1	MAIR_YEAST
22	64	11.3	708	1	TRAB_HUMAN
23	63.5	11.3	681	1	CAO2_RABIT
24	63.5	11.3	1294	1	YAB3_SCHPO
25	63	11.2	264	1	PSA4_DROME
26	63	11.2	1022	1	SCA4_RICPR
27	62.5	11.1	91	1	UTER_RABIT
28	62.5	11.1	223	1	YMA2_YEAST
29	62.5	11.1	610	1	VG12_BPT4
30	62	11.0	111	1	PSC1_RAT
31	62	11.0	938	1	V120_HSV7J
32	61.5	10.9	570	1	HEMI_KTULA
33	61.5	10.9	734	1	METE_THIEMA

34	61.5	10.9	2190	1	CCAD_CHICK
35	61	10.8	63	1	YEBW_ECOLI
36	61	10.8	142	1	GRPA_VACCC
37	61	10.8	605	1	VE1_HPV14
38	61	10.8	732	1	VE1_HPV5H
39	61	10.8	606	1	K086_MOUSE
40	61	10.8	830	1	JIP2_MOUSE
41	61	10.8	874	1	RPOL_BPSF6
42	61	10.8	1272	1	Y228_MRTJA
43	61	10.8	1966	1	CCAF_HUMAN
44	60.5	10.7	120	1	RHPA_BUCAI
45	60.5	10.7	180	1	MIP2_MOUSE

ALIGNMENTS

RESULT	ID	Score	% Query Match	Length	ID	Description
1	MGBB_HUMAN	564	87.9	95	1	MGBB_HUMAN
2	MGBA_HUMAN	281	49.8	93	1	MGBA_HUMAN
3	PSC3_RAT	195.5	34.7	95	1	PSC3_RAT
4	YRS9_MYCTU	71.5	12.7	131	1	YRS9_MYCTU
5	FTSY_RICPR	70	12.4	303	1	FTSY_RICPR
6	PSC2_RAT	69	12.2	112	1	PSC2_RAT
7	TC10_YEAST	68.5	12.1	684	1	TC10_YEAST
8	PR16_YEAST	68	12.1	1071	1	PR16_YEAST
9	TRPC_ACTCA	67.5	12.0	268	1	TRPC_ACTCA
10	SC10_YEAST	67.5	12.0	871	1	SC10_YEAST
11	RADC_HAEIN	66.5	11.8	234	1	RADC_HAEIN
12	TAZ1_SCHPO	66.5	11.8	663	1	TAZ1_SCHPO
13	IL15_MOUSE	65.5	11.6	162	1	IL15_MOUSE
14	VEL_HPV08	65.5	11.6	603	1	VEL_HPV08
15	G1T1_YEAST	65	11.5	2144	1	G1T1_YEAST
16	MTH_CAIAT	64.5	11.4	113	1	MTH_CAIAT
17	Y156_UREPA	64.5	11.4	153	1	Y156_UREPA
18	VE1_HPV21	64.5	11.4	603	1	VE1_HPV21
19	DYNL_ARATH	64.5	11.4	611	1	DYNL_ARATH
20	MAIR_YEAST	64	11.3	473	1	MAIR_YEAST
21	TRAB_HUMAN	64	11.3	708	1	TRAB_HUMAN
22	CAO2_RABIT	63.5	11.3	681	1	CAO2_RABIT
23	YAB3_SCHPO	63.5	11.3	1294	1	YAB3_SCHPO
24	PSA4_DROME	63	11.2	264	1	PSA4_DROME
25	SCA4_RICPR	63	11.2	1022	1	SCA4_RICPR
26	UTER_RABIT	62.5	11.1	91	1	UTER_RABIT
27	YMA2_YEAST	62.5	11.1	223	1	YMA2_YEAST
28	VG12_BPT4	62.5	11.1	610	1	VG12_BPT4
29	PSC1_RAT	62	11.0	111	1	PSC1_RAT
30	V120_HSV7J	62	11.0	938	1	V120_HSV7J
31	HEMI_KTULA	61.5	10.9	570	1	HEMI_KTULA
32	METE_THIEMA	61.5	10.9	734	1	METE_THIEMA
33	CCAD_CHICK	61.5	10.9	2190	1	CCAD_CHICK
34	YEBW_ECOLI	61	10.8	63	1	YEBW_ECOLI
35	GRPA_VACCC	61	10.8	142	1	GRPA_VACCC
36	VE1_HPV14	61	10.8	605	1	VE1_HPV14
37	VE1_HPV5H	61	10.8	732	1	VE1_HPV5H
38	K086_MOUSE	61	10.8	830	1	K086_MOUSE
39	JIP2_MOUSE	61	10.8	874	1	JIP2_MOUSE
40	RPOL_BPSF6	61	10.8	1272	1	RPOL_BPSF6
41	Y228_MRTJA	61	10.8	1966	1	Y228_MRTJA
42	CCAF_HUMAN	61	10.8	120	1	CCAF_HUMAN
43	RHPA_BUCAI	60.5	10.7	180	1	RHPA_BUCAI
44	MIP2_MOUSE	60.5	10.7	180	1	MIP2_MOUSE
45	MIP2_MOUSE	60.5	10.7	180	1	MIP2_MOUSE

RESULT	ID	Score	% Query Match	Length	ID	Description
1	MGBB_HUMAN	564	87.9	95	1	MGBB_HUMAN
2	MGBA_HUMAN	281	49.8	93	1	MGBA_HUMAN
3	PSC3_RAT	195.5	34.7	95	1	PSC3_RAT
4	YRS9_MYCTU	71.5	12.7	131	1	YRS9_MYCTU
5	FTSY_RICPR	70	12.4	303	1	FTSY_RICPR
6	PSC2_RAT	69	12.2	112	1	PSC2_RAT
7	TC10_YEAST	68.5	12.1	684	1	TC10_YEAST
8	PR16_YEAST	68	12.1	1071	1	PR16_YEAST
9	TRPC_ACTCA	67.5	12.0	268	1	TRPC_ACTCA
10	SC10_YEAST	67.5	12.0	871	1	SC10_YEAST
11	RADC_HAEIN	66.5	11.8	234	1	RADC_HAEIN
12	TAZ1_SCHPO	66.5	11.8	663	1	TAZ1_SCHPO
13	IL15_MOUSE	65.5	11.6	162	1	IL15_MOUSE
14	VEL_HPV08	65.5	11.6	603	1	VEL_HPV08
15	G1T1_YEAST	65	11.5	2144	1	G1T1_YEAST
16	MTH_CAIAT	64.5	11.4	113	1	MTH_CAIAT
17	Y156_UREPA	64.5	11.4	153	1	Y156_UREPA
18	VE1_HPV21	64.5	11.4	603	1	VE1_HPV21
19	DYNL_ARATH	64.5	11.4	611	1	DYNL_ARATH
20	MAIR_YEAST	64	11.3	473	1	MAIR_YEAST
21	TRAB_HUMAN	64	11.3	708	1	TRAB_HUMAN
22	CAO2_RABIT	63.5	11.3	681	1	CAO2_RABIT
23	YAB3_SCHPO	63.5	11.3	1294	1	YAB3_SCHPO
24	PSA4_DROME	63	11.2	264	1	PSA4_DROME
25	SCA4_RICPR	63	11.2	1022	1	SCA4_RICPR
26	UTER_RABIT	62.5	11.1	91	1	UTER_RABIT
27	YMA2_YEAST	62.5	11.1	223	1	YMA2_YEAST
28	VG12_BPT4	62.5	11.1	610	1	VG12_BPT4
29	PSC1_RAT	62	11.0	111	1	PSC1_RAT
30	V120_HSV7J	62	11.0	938	1	V120_HSV7J
31	HEMI_KTULA	61.5	10.9	570	1	HEMI_KTULA
32	METE_THIEMA	61.5	10.9	734	1	METE_THIEMA
33	CCAD_CHICK	61.5	10.9	2190	1	CCAD_CHICK
34	YEBW_ECOLI	61	10.8	63	1	YEBW_ECOLI
35	GRPA_VACCC	61	10.8	142	1	GRPA_VACCC
36	VE1_HPV14	61	10.8	605	1	VE1_HPV14
37	VE1_HPV5H	61	10.8	732	1	VE1_HPV5H
38	K086_MOUSE	61	10.8	830	1	K086_MOUSE
39	JIP2_MOUSE	61	10.8	874	1	JIP2_MOUSE
40	RPOL_BPSF6	61	10.8	1272	1	RPOL_BPSF6
41	Y228_MRTJA	61	10.8	1966	1	Y228_MRTJA
42	CCAF_HUMAN	61	10.8	120	1	CCAF_HUMAN
43	RHPA_BUCAI	60.5	10.7	180	1	RHPA_BUCAI
44	MIP2_MOUSE	60.5	10.7	180	1	MIP2_MOUSE
45	MIP2_MOUSE	60.5	10.7	180	1	MIP2_MOUSE

RESULT	ID	Score	% Query Match	Length	ID	Description
1	MGBB_HUMAN	564	87.9	95	1	MGBB_HUMAN
2	MGBA_HUMAN	281	49.8	93	1	MGBA_HUMAN
3	PSC3_RAT	195.5	34.7	95	1	PSC3_RAT
4	YRS9_MYCTU	71.5	12.7	131	1	YRS9_MYCTU
5	FTSY_RICPR	70	12.4	303	1	FTSY_RICPR
6	PSC2_RAT	69	12.2	112	1	PSC2_RAT
7	TC10_YEAST	68.5	12.1	684	1	TC10_YEAST
8	PR16_YEAST	68	12.1	1071	1	PR16_YEAST
9	TRPC_ACTCA	67.5	12.0	268	1	TRPC_ACTCA
10	SC10_YEAST	67.5	12.0	871	1	SC10_YEAST
11	RADC_HAEIN	66.5	11.8	234	1	RADC_HAEIN
12	TAZ1_SCHPO	66.5	11.8	663	1	TAZ1_SCHPO
13	IL15_MOUSE	65.5	11.6	162	1	IL15_MOUSE
14	VEL_HPV08	65.5	11.6	603	1	VEL_HPV08
15	G1T1_YEAST	65	11.5	2144	1	G1T1_YEAST
16	MTH_CAIAT	64.5	11.4	113	1	MTH_CAIAT
17	Y156_UREPA	64.5	11.4	153	1	Y156_UREPA
18	VE1_HPV21	64.5	11.4	603	1	VE1_HPV21
19	DYNL_ARATH	64.5	11.4	611	1	DYNL_ARATH
20	MAIR_YEAST	64	11.3	473	1	MAIR_YEAST
21	TRAB_HUMAN	64	11.3	708	1	TRAB_HUMAN
22	CAO2_RABIT	63.5	11.3	681	1	CAO2_RABIT
23	YAB3_SCHPO	63.5	11.3	1294	1	YAB3_SCHPO
24	PSA4_DROME	63	11.2	264	1	PSA4_DROME
25	SCA4_RICPR	63	11.2	1022	1	SCA4_RICPR
26	UTER_RABIT	62.5	11.1	91	1	UTER_RABIT
27	YMA2_YEAST	62.5	11.1	223	1	YMA2_YEAST
28	VG12_BPT4	62.5	11.1	610	1	VG12_BPT4
29	PSC1_RAT	62	11.0	111	1	PSC1_RAT
30	V120_HSV7J	62	11.0	938	1	V120_HSV7J
31	HEMI_KTULA	61.5	10.9	570	1	HEMI_KTULA
32	METE_THIEMA	61.5	10.9	734	1	METE_THIEMA
33	CCAD_CHICK	61.5	10.9	2190	1	CCAD_CHICK
34	YEBW_ECOLI	61	10.8	63	1	YEBW_ECOLI
35	GRPA_VACCC	61	10.8	142	1	GRPA_VACCC
36	VE1_HPV14	61	10.8	605	1	VE1_HPV14
37	VE1_HPV5H	61	10.8	732	1	VE1_HPV5H
38	K086_MOUSE	61	10.8	830	1	K086_MOUSE
39	JIP2_MOUSE	61	10.8	874	1	JIP2_MOUSE
40	RPOL_BPSF6	61	10.8	1272	1	RPOL_BPSF6
41	Y228_MRTJA	61	10.8	1966	1	Y228_MRTJA
42	CCAF_HUMAN	61	10.8	120	1	CCAF_HUMAN
43	RHPA_BUCAI	60.5	10.7	180	1	RHP

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF071219; AAC79996.1; -
 CC EMBL: AJ224173; CA11865.1; -
 CC MIM: 604398; -
 CC InterPro: IPR003627; Mammaglobin_prostin.
 CC InterPro: IPR000329; Uterogloblin.
 CC Pfam: PF01099; Uterogloblin; 1.
 CC ProDom: PD029354; Mammaglobin_prostin; 1.
 CC PROSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
 CC PROSITE: PS00404; UTEROGLOBIN_2; FALSE_NEG.
 CC Signal: Glycoprotein.
 CC SIGNAL 1 18 MAMMAGLOBIN B.
 CC CHAIN 19 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 95 AA: 10884 MW: 0719738289f89f8d CRC64;

Query Match 87.9%; Score 496; DB 1; Length 95;
 Best Local Similarity 100.0%; Pred. No. 1.2e-43;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 MKLWVLMALALHCHYADSGCKLEDMVEKTSIPYKELLOEFDSDAAAEAMG 73
 DB 1 MKLWVLMALALHCHYADSGCKLEDMVEKTSIPYKELLOEFDSDAAAEAMG 60
 OY 74 KFKOCFLNOSHRTLNKFGIMHTVYDSIWCNMKN 108
 DB 61 KFKOCFLNOSHRTLNKFGIMHTVYDSIWCNMKN 95

RESULT 2

MGBA_HUMAN STANDARD; PRT; 93 AA.
 AC 013396;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mammaglobin A precursor (Mammaglobin 1).
 DE MGB1 OR UGB2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=breast;
 RX MEDLINE=96223698; PubMed=8631025;
 RA Watson M.A., Fleming T.P.;
 RT "Mammaglobin, a mammary-specific member of the uterogloblin gene family, is overexpressed in human breast cancer.";
 RL Cancer Res. 56:860-865(1996).
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98147371; PubMed=9488047;
 RA Watson M.A., Darlow C., Zimonjic D.B., Popescu N.C., Fleming T.;
 RT "Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uterogloblin gene family localized to chromosome 11q13.";
 RL Oncogene 16:817-824(1998).
 CC -1- TISSUE SPECIFICITY: MAMMARY-SPECIFIC. OVER-EXPRESSED IN BREAST CANCER.
 CC -1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U3147; AAC50375.1; -
 CC EMBL: AF015224; AAC39608.1; -
 CC MIM: 605562; -
 CC InterPro: IPR003627; Mammaglobin_prostin.
 CC InterPro: IPR000329; uterogloblin.
 CC Pfam: PF01099; Uterogloblin; 1.
 CC ProDom: PD029354; Mammaglobin_prostin; 1.
 CC SMART: SM00096; UTR; 1.
 CC PROSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
 CC PROSITE: PS00404; UTEROGLOBIN_2; 1.
 CC Signal: Glycoprotein.
 CC SIGNAL 1 18 POTENTIAL.
 CC CHAIN 19 93 MAMMAGLOBIN A.
 CC CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 93 AA: 10499 MW: 28968C43BF053E2 CRC64;

Query Match 49.8%; Score 281; DB 1; Length 93;
 Best Local Similarity 58.7%; Pred. No. 7.1e-22;
 Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

OY 14 MKLWVLMALALHCHYADSGCKLEDMVEKTSIPYKELLOEFDSDAAAEAMG 73
 DB 1 MKLWVLMALALHCHYADSGCKLEDMVEKTSIPYKELLOEFDSDAAAEAMG 60
 OY 74 KFKOCFLNOSHRTLNKFGIMHTVYDSIWCNMKN 105
 DB 61 KFKOCFLNOSHRTLNKFGIMHTVYDSIWCNMKN 92

RESULT 3

PSC3_RAT STANDARD; PRT; 95 AA.
 AC P02780; 063463;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Prostatic steroid-binding protein C3 chain precursor (Prostatein peptide C3).
 DE Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83082848; PubMed=6294095;
 RA Parker M.G., White R., Hurst H., Needham M., Tilly R.;
 RT "Prostatic steroid-binding protein. Isolation and characterization of C3 genes.";
 RL J. Biol. Chem. 258:12-15(1983).
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83238526; PubMed=6190812;
 RA Viskochil D.H., Perry S.T., Lea O.A., Stafford D.W., Wilson E.M., French F.S.;
 RT "Isolation of two genomic sequences encoding the Mr = 14,000 subunit of rat prostatein.";
 RL J. Biol. Chem. 258:8861-8866(1983).
 RN 13
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92165796; PubMed=1537831;
 RA Tan J.A., Marschke K.B., Ho K.-C., Perry S.T., Wilson E.M., French F.S.;
 RT "Response elements of the androgen-regulated C3 gene.";
 RL J. Biol. Chem. 267:4456-4466(1992).
 RN 14


```

ERRATUM=92218467; PubMed=1339454;
RX MEDLINE=81188769; PubMed=7014218;
RA Tan J.A., Marschke K.B., Ho K.C., Perry S.T., Wilson E.M.,
RA French F.S.,
RL J. Biol. Chem. 267:7958-7958(1992).
RN 151
RP SEQUENCE OF 19-95.
RX MEDLINE=81188769; PubMed=7014218;
RA Peeters B., Rombaux W., Mous J., Heyns W.;
RT "Structural studies on rat prostatic binding protein. The primary
RT structure of its glycosylated component C3.";
RL Eur. J. Biochem. 115:115-121(1981).
CC -1- FUNCTION: PART OF PROSTATEIN WHICH IS THE MAJOR SECRETORY
CC GLYCOPROTEIN OF VENTRAL PROSTATE GLAND. STEROID-BINDING PROTEIN;
CC CAN BIND NON-POLAR STEROIDS, CHOLESTEROL AND A GROUP OF SMALL
CC PROLINE-RICH PEPTIDES.
CC -1- SUBUNIT: PROSTATEIN IS COMPOSED OF THREE DIFFERENT PEPTIDES CALLED
CC C1, C2 AND C3. THESE FORM COVALENT C1:C3 (P) AND C2:C3 (S)
CC HETERODIMERS WHOSE NONCOVALENT ASSOCIATION FORMS TETRAMERIC
CC (C1:C3/C3:C2) PROSTATEIN MOLECULES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: VENTRAL PROSTATE.
CC -1- INDUCTION: ANDROGEN DEPENDENT, AS SHOWN BY THE DECREASE IN THE
CC LEVEL OF THE PROTEIN FOLLOWING CASTRATION.
CC -1- MISCELLANEOUS: C3 IS ENCODED BY TWO DIFFERENT GENES.
CC -1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPONHILIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: V01263; CAA24577.1; -.
CC EMBL: M71245; AAA41965.1; -.
CC PIR: A03250; B0R73.
CC InterPro: IPR003627; Mammagln.prostin.
CC InterPro: IPR000329; Uteroglobln.
CC DR Pfam: PF01099; Mammaglobln.1.
CC DR ProDom: PD029354; Mammaglobln.prostin.1.
CC DR PROSITE: PS00403; UTEROGLOBLIN.1; FALSE_NRG.
CC DR PROSITE: PS00404; UTEROGLOBLIN.2; 1.
CC KM Signal: glycoprotein; Steroid-binding.
CC FT SIGNAL 1 18 PROSTATIC STEROID-BINDING PROTEIN C3
CC FT CHAIN 19 95 CHAIN
CC FT CARBOHYD 35 35 N-LINKED (GLCNAC. .).
CC FT CONFLICT 53 53 D -> A (IN REF. 3).
CC FT CONFLICT 79 79 G -> S (IN REF. 2).
CC FT SEQUENCE 95 AA: 10730 MW: 17571.00C882E375 CRC64:
CC
Query Match 34.7%: Score 195.5; DB 1; Length 95;
Best Local Similarity 35.8%; Pred. No. 3; E-13;
Matches 34; Conservative 30; Mismatches 30; Indels 1; Gaps 1
OY 14 MKLVAVTALALLLHGYAD-SGCKLDEWVEKTIINDSISPEYKELQLEFIDSDAAAEAM 72
DB 1 MKLVFLFLVLIPIICVYASGSGCSILDEVIKGTINVTVLADWKLKVPYVODHFTFKAV 60
OY 73 GKFKOCFLNOSHRLTXNGFLMMHIVYOSINCMKMS 107
DB 61 KFKOCFLDQTDKLTLENGVMHEAIFNSSCQPS 95
RESULT 4
ID YR59_MYCTU STANDARD; PRT; 131 AA.
AC 033301;
DT 30-MAY-2000 (Rel. 39, Created)

```

```

DN 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DT Hypothetical 14.4 kDa protein Rv2759C.
GN Rv2759c OR MT829 OR MTV002.24C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCPI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekle F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Dellin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Oliver S., Osborne J., Quail M.A., McLean J., Moule S., Murphy L.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri R., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/Genbank/DBDJ databases.
CC 1- SIMILARITY: BELONGS TO THE UP0010 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL008967; CAA15555.1; -
DR EMBL: AE007110; AK647148.1; -
DR TIGR: MT829; -
DR "tubercu1ist; kw2759c; -
KW Hypothetical protein; Complete proteome.
KW SEQUENCE 131 AA; 14372 MW; 9C12AP0AC674065B CRC64;
SQ

Query Match 12.7% Score 71.5; DB 1; Length 131;
Best local similarity 21.2% Pval. No. 1.7;
Matches 21; Conservative 18; Mismatches 35; Indels 25; Gaps 13

QY 18 WYMAALLLCYDMSGCKLEDWEKTIINDSISPEYKELQ----- 60
DB 1 MIVDTSAIVAVSGESGQVYKALELSPNRMSPAPVYVELCAIMQRRDPEISRLVDRL 60
QY 61 -----ETIDSDAAAEAMCKEFQCFPLNOSHRTLKAFG 91
DB 61 IDDYGIQVEAVDDAQARVAQAQAYRDYGRGSGIPARLNLG 99

RESULT 5
FSTY_RICPR STANDARD; PRT; 303 AA.
AC 005948;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell division protein fsty homolog.

```

```

DN 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DT Hypothetical 14.4 kDa protein Rv2759C.
GN Rv2759c OR MT829 OR MTV002.24C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCPI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekle F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Dellin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Oliver S., Osborne J., Quail M.A., McLean J., Moule S., Murphy L.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri R., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: BELONGS TO THE UP0010 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL008967; CAA15555.1; -
DR EMBL: AE007110; AAK47148.1; -
DR TIGR: MT829; -
DR "tubercu1ist; kw2759c; -
KW Hypothetical protein; Complete proteome.
KW SEQUENCE 131 AA; 14372 MW; 9C12AP0AC674065B CRC64;
SQ

```


FT CONFLICT 26 26 MISSING (IN REF. 3).
 FT CONFLICT 88 88 I -> T (IN REF. 3).
 FT CONFLICT 96 112 VWLOINPGRWFESEIN -> YGYK (IN REF. 3).
 SQ SEQUENCE 112 AA: 12828 MW: DA65A6A82E677864 CRC64;

Query Match 12.2% Score 69; DB 1; Length 112;
 Best Local Similarity 27.4%; Pred. No. 2.5;
 Matches 29; Conservative 21; Mismatches 34; Indels 22; Gaps 6;

OY 11 RLAKMLVLMALLLHCVADSG-----CKLEDM-VEKTIINSDISPEKELQEF 62
 DB 2 RLSTLTFLVVC-----CLEANGQLAGVCCALDVTITFLINE-----ELKRELEEF 53

OY 63 IDSDAAEAMGKFKOCFLNQSHRTLNFGIMHT--VYDSICNMK 106
 DB 54 DAPPEAVEAMIKVKRCT-----NKIMYCDRLSMGTSILVIMLKCDVK 95

RESULT 7

TC10_YEAST STANDARD; PRT; 684 AA.
 AC P50273;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE TCM10 protein.
 DE TCM10 OR YDR350C OR D9476.9.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;

SEQUENCE FROM N.A.
 SPRAIN-MH125;

RA Zhang Y., Robinson K.M., Lemire B.D.;
 RA submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Pavesio A., Fulton L., Gattling S., Greco T., Kirsten J.,
 RA Kuehba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Madis E., Meneses S., Miller N., Nuan H., Paley A., Peluso D.,
 RA Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
 RA submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U32306; AAA74031.1; -;
 DR EMBL: U28372; AA64786.1; -;
 DR SGD: S0002758; TCM10.
 FT CONFLICT 592 684 GARSYNKILEGGFEIRIMALIOIKDGMFPKPNDETL
 FT NIIRETKSLN -> EHAFGTIRYSSGALKSGIWL (IN
 FT REF. 2).
 FT SEQUENCE 684 AA: 79755 MW: A88992848F8F9A4 CRC64;

Query Match 12.1% Score 68.5; DB 1; Length 684;
 Best Local Similarity 23.0%; Pred. No. 20;
 Matches 23; Conservative 22; Mismatches 44; Indels 11; Gaps 3;

OY 1 CHARLNTDSRLAMKMLVLMALLLHCVAD--SGCKLEDMVEKT-----INSDI 50
 DB 385 CKYLLSSYSKASANTLMDGIMGWLHYKSKTTLTAPTLARELKNNNILPMTMRIGSPV 444

OY 51 SIPEKELQEFISDDAAEAMGKFKOCFLNQSHRTLNK 90
 DB 445 TVPILTELRYSLSSSVLSLESGQFKNCLDI.YTK-YKSF 483

RESULT 8

PR16_YEAST STANDARD; PRT; 1071 AA.
 AC P15938;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE pre-mRNA splicing factor RNA helicase PRP16.
 GN PRP16 OR YKR086W OR YKR406.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;

SEQUENCE FROM N.A.
 MEDLINE=90182651; PubMed=2138057;
 RA Burgess S., Couto J.R., Guthrie C.;
 RT "A putative ATP binding protein influences the fidelity of
 RT branchpoint recognition in yeast splicing.";
 RL Cell 60:705-717(1990).

SEQUENCE FROM N.A.
 MEDLINE=94262327; PubMed=8203164;
 RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
 RA Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
 RT "The complete sequence of an 18,002 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
 RT and six new open reading frames.";
 RL Yeast 10:231-245(1994).

FUNCTION.
 RP MEDLINE=91125476; PubMed=1825134;
 RA Schwer B., Guthrie C.;

RT "PRP16 is an RNA-dependent ATPase that interacts transiently with the
 RT spliceosome.";
 RL Nature 349:494-499(1991).

CC -1- FUNCTION: INFLUENCES THE FIDELITY OF BRANCHPOINT RECOGNITION IN
 CC YEAST SPLICING. THIS IS RNA-DEPENDENT ATPASE WHICH IS ESSENTIAL
 CC FOR VIABILITY. IT MAY MEDIATE ONE OF THE MANY ATP-REQUIRING STEPS
 CC OF SPLICOSOME ASSEMBLY AND THAT ACCURACY OF BRANCHPOINT
 CC RECOGNITION MAY BE COUPLED TO ATP BINDING AND/OR HYDROLYSIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. BINDS TO THE SPLICOSOME
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH
 CC SUBFAMILY. PRP16 ORTHOLOG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M31524; AAA34911.1; -;
 DR EMBL: Z27116; CA81637.1; -;
 DR EMBL: Z28311; CA82165.1; -;
 DR PIR: S38164; S38164.
 DR PIR: S39127; S39127.
 DR SGD: S0001794; PRP16.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAH_ATP_helicase.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; helicase_C.1.
 DR SMART: SM00487; DEAD_C.1.
 DR SMART: SM00490; HELIC_C.1.
 DR PROSITE: PS00690; DEAH_ATP_HELICASE.1.
 KW mRNA processing; mRNA splicing; Helicase; ATP-binding;

KW Nuclear protein.
 FT NP_BIND 373 380 ATP (BY SIMILARITY).
 FT SITE 473 476 DEAH BOX.
 FT MUTAGEN 386 386 Y->D: SUPPRESSOR PHENOTYPE.
 FT CONFLICT 698 698 A->R (IN REF. 1).
 SQ SEQUENCE 1071 AA: 121652 MW: 87D4C4A18F5DAE1 CRC64;

Query Match 12.1%; Score 68; DB 1; Length 1071;
 Best Local Similarity 25.7%; Pred. No. 37;
 Matches 26; Conservative 19; Mismatches 48; Indels 8; Gaps 3;

QY 7 TDSRLAMKLM-VLMIALLLHCYADSGCKLLEDMVEKINSISIPYKELLQ---- 60
 DB 439 TDSCTKLVYTGILRELLDDTDKSCVYLDEAHERSLMTDILGFFKLLARRD 498
 QY 61 -EFIDSDAAEAMGKFCFLNOSHRTKFMGLMHTYDS 100
 DB 499 LKLIITSATMNAK-KESAFPGNAPQFTIIPGRTPPYOTIYTS 538

RESULT 9
 TRPC_AC1CA STANDARD; PRT; 268 AA.
 ID TRPC_AC1CA
 AC P00911;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Indole-3-glycerol phosphate synthase (EC 4.1.1.48) (IGPS).
 OS TRPC.
 OS Actinobacter calcoaceticus.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 OC Actinobacter.
 NC NCB1_TaxID=471;
 RN [1]
 RP MEDLINE=88174326; PubMed=6599977;
 RA Kaplan J.B., Goncharoff P., Seibold A.M., Nichols B.P.;
 RT "Nucleotide sequence of the Actinobacter calcoaceticus trpGDC gene cluster."
 RL Mol. Biol. Evol. 14:456-472(1984).
 CC -1- CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-1'-deoxy-D-ribose 5-phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.
 CC -1- PATHWAY: FOURTH STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
 CC -1- SIMILARITY: BELONGS TO THE TRPC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M36636; AAA21905.1; -
 CC PIR: A01133; GMECC.
 CC DR HSSE: P00909; IPII.
 CC DR InterPro: IPR001468; IGPS.
 CC DR Pfam: PF00218; IGPS; 1.
 CC DR Prodom: PD001511; IGPS; 1.
 CC DR PROSITE: PS00614; IGPS; 1.
 CC KW Tryptophan biosynthesis; lyase; Decarboxylase.
 CC SQ SEQUENCE 268 AA: 30216 MW: 8DBAC0505D5A7527 CRC64;

Query Match 12.0%; Score 67.5; DB 1; Length 268;
 Best Local Similarity 26.4%; Pred. No. 9.3;
 Matches 19; Conservative 16; Mismatches 28; Indels 9; Gaps 2;

QY 25 LLHCYADSGCKLLEDMVEKINSISIPYKELLQEFIDSDAAEAMGKFCFLNOSH 84
 DB 143 LIVACLSLDD--QOLEEMSKTAFEDYL-----DVLVEVHDEQLERALKLSQCLIGVNN 193

QY 85 RFLKNFGIMMHT 96
 DB 194 RNLKTFVDVLTMT 205

RESULT 10
 SC10_YEAST STANDARD; PRT; 871 AA.
 ID SC10_YEAST
 AC Q06245; P87329;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exocyst complex component SEC10.
 GN SEC10 OR YLR166C OR L9362.12.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NC NCB1_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 347-353 AND 856-867.
 RX MEDLINE=97133278; PubMed=8978675;
 RA Terbush D.R., Maurice T., Routh D., Novick P.;
 RT "The Exocyst is a multiprotein complex required for exocytosis in Saccharomyces cerevisiae."
 RL EMBO J. 15:6483-6494(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Javellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
 RA Taich A., Trevasakis E., Vignati D., Wilcox L., Woldman P., Vaudin M.,
 RA Wilson R., Waterston R.;
 RL Submitted (Mar-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR EXOCYTOSIS.
 CC -1- SUBUNIT: SEC3, SEC6, SEC8, SEC10, SEC15 AND EXO70 ARE COMPONENT OF EXOCYST COMPLEX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y08789; CAA70041.1; -
 CC DR EMBL: U51921; AAB67490.1; -
 CC DR EMBL: U17246; AAB67476.2; -
 CC DR SGD: S0004156; SEC10.
 CC KW Transport; Protein transport; Golgi stack; Coiled coil.
 CC FT DOMAIN 74 101 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 871 AA: 100341 MW: A32073099815388D CRC64;

Query Match 12.0%; Score 67.5; DB 1; Length 871;
 Best Local Similarity 28.4%; Pred. No. 33;
 Matches 25; Conservative 17; Mismatches 29; Indels 17; Gaps 4;

QY 4 RLNTSSRLAMKLMVLMIALLLHCYADSGCKLLEDMVEKINSISIPYKELLQ-EF 62
 DB 190 KLNKLSVKL-MKNLLIT-----SSKLETSSIPKTIINTKLVIEKSEMMENEL 235

QY 63 IDSDAEAMGKFCFLNOSHRTKFMGLMHTYDS 90
 DB 236 LENFNSATVRENNFTK--LNEIATILLNNF 261

RESULT 11
 RADCL_HAEIN STANDARD; PRT; 234 AA.
 ID RADCL_HAEIN

```

AC P44952;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA repair protein radc homolog.
GN RADC OR H10952.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=93350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McEnany K., Sutton G., Fitzhugh W., Field C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uettermack T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: INVOLVED IN DNA REPAIR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RADC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32776; AAC22613.1;
CC DR TIGR: H10952;
CC DR InterPro: IPR001405; RadC.
CC DR ProDom: PD007415; RadC; 1.
CC DR PROSITE: PS01302; RADC; 1.
CC DR DNA repair; Complete proteome.
CC KW SEQUENCE 234 AA; 26774 MW; 25472EB289E5DA69 CRC64;
SQ

Query Match 11.8%; Score 66.5; DB 1; Length 234;
Best Local Similarity 26.6%; Pred. No. 10;
Matches 21; Conservative 15; Mismatches 26; Indels 17; Gaps 3;

OY 37 LLEDNVEKTIINSIPYKELLOEFIDSDAAEAMGKFCFLNOSHRLTKNFGIMHT 96
DB 103 LKQDMSTPIIND--PEYKL--FLTELOHEEREVWVFLDNQHLIKERFLDT 156
OY 97 VY-----DSIWCN 104
DB 157 IYSAVYPREIKKALYCN 175

RESULT 12
TAZI_SCHPO STANDARD; PRT; 663 AA.
ID TAZI_SCHPO
AC P79005; 000049;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Telomere length regulator taz1.
GN TAZ1 OR MYB OR MYB OR SPAC16A10.07C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
NCBI_TaxID=4896;
SQ

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=97186578; PubMed=9034194;
RA Cooper J.P., Nimmo E.R., Allshire R.C., Cech T.R.;
RT "Regulation of telomere length and function by a Myb-domain protein
RT in fission yeast."
RL Nature 385:744-747(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: REGULATES TELOMERE LENGTH AND FUNCTION. REQUIRED FOR THE
CC REPRESSION OF TELOMERE-ADJACENT GENE EXPRESSION AND FOR NORMAL
CC MEIOSIS OR SPORULATION. IT MAY BE A NEGATIVE REGULATOR OF THE
CC TELOMERE-REPLICATING ENZYME, TELOMERASE, OR MAY PROTECT AGAINST
CC ACTIVATION OF TELOMERASE-INDEPENDENT PATHWAYS OF TELOMERE
CC ELONGATION. IT MAY BE INVOLVED IN THE INTERACTIONS BETWEEN
CC CHROMOSOMES AND SPINDLE PROTEINS, DISRUPTION OF THESE
CC INTERACTIONS WOULD LEAD TO DEFECTIVE MEIOSIS.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y09406; CAJ70568.1;
CC DR EMBL: Z97185; CAB10000.1;
CC DR InterPro: IPR001005; MYB-DNA_bind.
CC DR Pfam: PF00249; myb_DNA-binding; 1.
CC DR SMART: SM00395; SANT; 1.
CC DR PROSITE: PS00037; MYB-1; 1.
CC DR PROSITE: PS00334; MYB-2; FALSE NEG.
CC DR PROSITE: PS50090; MYB-3; FALSE NEG.
CC DR Telomere; Nuclear protein; DNA-binding; Repeat.
CC KW DNA_BIND 556 612 MYB 74647 MW; 5852P21518031152 CRC64;
CC FT SEQUENCE 663 AA; 74647 MW; 5852P21518031152 CRC64;
SQ

Query Match 11.8%; Score 66.5; DB 1; Length 663;
Best Local Similarity 21.0%; Pred. No. 31;
Matches 22; Conservative 28; Mismatches 38; Indels 17; Gaps 5;

OY 8 DSSRLAKLIMVIMIALLLHCYADSGCKLHIDVEKTIINSIDISIPYKELLOEFIDSDA 67
DB 119 DONMAIVRSIMDIKASLV-----NDCONTANNAEOKVMVMSAI--FSSESKDIYNPES 171
OY 68 AAEAMGK--FPCFLNOSHRLTKNFGIMHTV-----YDSINC 103
DB 172 FSRIGKRETVADYFENQOLITK-YGLEFRTLPFSYIKKDYAVYC 215

RESULT 13
IL15_MOUSE STANDARD; PRT; 162 AA.
ID IL15_MOUSE
AC P48346;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Interleukin-15 precursor (IL-15).
GN IL15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
SQ

```

RC STRAIN-WC/REI X C57BL/6J; TISSUE-Bone marrow;
 RA MEDLINE-95278940; Pubmed-7759105;
 RA Anderson D.M., Johnson L., Giaccum M.B., Copeland N.G., Gilbert D.J.,
 RA Jenkins N.A., Valentine V., Kirstein M.N., Shapiro D.N., Morris S.W.,
 RA Grabstein K., Cosman D.;
 RT *Chromosomal assignment and genomic structure of IL15.*;
 CC Genomics 25:701-706(1995).
 RL -----
 CC -1- FUNCTION: CYTOKINE THAT STIMULATES THE PROLIFERATION OF T-
 CC LYMPHOCYTES. STIMULATION BY IL-15 REQUIRES INTERACTION OF IL-15
 CC WITH COMPONENTS OF IL-2R, INCLUDING IL-2R BETA AND PROBABLY IL-2R
 CC GAMMA BUT NOT IL-2R ALPHA.
 CC -1- SIMILARITY: BELONGS TO THE IL-15 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U14332; AAA5377.1; -.
 DR MGD: MGI:103014; IL15.
 DR InterPro: IPR003443; Interleukin_15.
 DR Pfam: PF02372; IL15; 1.
 DR Cytokine; glycoprotein; Signal.
 KW SIGNAL 1 29
 FT PROPEP 30 48 POTENTIAL.
 FT CHAIN 49 162 INTERLEUKIN-15.
 FT DISULFID 83 133 POTENTIAL.
 FT DISULFID 90 136 POTENTIAL.
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 162 AA; 18593 MW; 68C971496CBE296 CRC64;
 QY 31 ADSGCKLEDVVEKTSINSDISIPYKELLQEFIDSDAAAEAMGKRCQCLNS 83
 DB 129 AESGCKCEKEELFEKFT-----PFEPLQSF-----RIVQMPINFS 162
 Query Match 11.6%; Score 65.5; DB 1; Length 162;
 Best Local Similarity 34.0%; Pred. No. 8.6;
 Matches 18; Conservative 6; Mismatches 10; Indels 19; Gaps 2;
 RESULT 14
 VE1_HPV08 STANDARD: PRT; 603 AA.
 ID VE1_HPV08
 AC P06420;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Replication protein E1.
 GN E1.
 OS Human papillomavirus type 8.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomaviruses.
 OX NCBI_TaxID=10579;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-86200410; Pubmed-3009874;
 RA Fuchs P.G., Iftner T., Weninger J., Pfister H.,
 RT "Papillomaviridae verruiformis-associated human papillomavirus 8:
 RT genomic sequence and comparative analysis.";
 RL J. Virol. 58:626-634(1986).
 CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
 CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
 CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
 CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M12737; -; NOT_ANNOTATED_CDS.
 DR PIR: A03656; M1WLB.
 DR InterPro: IPR001177; Papillom_E1.
 DR Pfam: PF00519; E1; 1.
 DR Pfam: PF00524; ELN; 1.
 KW Early protein: DNA replication; Helicase; ATP-binding; DNA-binding;
 FT Nuclear protein.
 FT NUC_BIND 431 438 ATP (POTENTIAL).
 SQ SEQUENCE 603 AA; 68821 MW; 0813860098DA8A8D CRC64;
 QY 32 DSGCKLEDVVEKTSINSDISIP-----EYKELLQEFIDSDAAAEAMGK 74
 DB 112 DSGVELTLNNEADVDYSEVEPAIDSRPEDEGGSGALDIDYTLALR--SSMTKATLMAK 168
 QY 75 FKQCFLN-----QSHRTLKNFGILM-----MRTVYDS 100
 DB 169 FKFAFGDGFNELLRQFQSKYTCGNYWVAAYVAHDYVES 207
 Query Match 11.6%; Score 65.5; DB 1; Length 603;
 Best Local Similarity 22.2%; Pred. No. 36;
 Matches 22; Conservative 21; Mismatches 23; Indels 33; Gaps 4;
 RESULT 15
 GLUT1_YEAST STANDARD: PRT; 2144 AA.
 ID GLUT1_YEAST
 AC Q12680; Q12290;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Glutamate synthase [NADPH] precursor (EC 1.4.1.13) (NADPH-GOGAT).
 GN GLUT1 OR YDL171C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX STRAIN-CN36;
 RX MEDLINE-97082505; Pubmed-8923741;
 RA Filletici P., Martegani M.P., Valenzuela L., Gonzalez A., Ballario P.;
 RT "Sequence of the GLUT1 gene from Saccharomyces cerevisiae reveals the
 RT domain structure of yeast glutamate synthase.";
 RL Yeast 12:1359-1366(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-S288C;
 RA Pohl T.M.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 2 L-glutamate + NADP(+) = L-glutamine + 2-
 CC OXOGlutamate + NADPH;
 CC -1- COFACTOR: BINDS A 3FE-4S CLUSTER: FAD AND FMN.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X89221; CAA61505.1; -.

Search completed: June 20, 2002, 11:06:16
Job time: 212 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: June 20, 2002, 11:02:09 ; Search time 25.21 Seconds
(without alignments)
741.113 Million cell updates/sec

Title:	US-09-673-395A-238
Perfect score:	564
Sequence:	1 CHARLNTDSSRLAKMLMVL.....NFGLMHTYVDSIWCNKNKS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```

Searched:      562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters:  562222

```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```

1:  sp.archaea:19:*
2:  sp.bacteria:*
3:  sp.fungi:*
4:  sp.human:*
5:  sp.invertebrate:*
6:  sp.mammal:*
7:  sp.mhc:*
8:  sp.organelle:*
9:  sp.phage:*
10: sp.plant:*
11: sp.podant:*
12: sp.virus:*
13: sp.vetelbrate:*
14: sp.unclassified:*
15: sp_virus:*
16: sp_bacteria:*
17: sp_archaea:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	283	50.2	93	6	09GK63	09GK63 oryctolagus
2	232	41.1	93	6	09GK61	09GK61 oryctolagus
3	229	40.6	93	6	09GK64	09GK64 oryctolagus
4	227	40.2	93	6	09GK62	09GK62 oryctolagus
5	191.5	34.0	95	11	09JHB9	09JHB9 ratius norv
5	191.5	34.0	95	11	09JHB9	09JHB9 mesocricet
6	178.5	31.6	94	11	09QXF2	09QXF2 mesocricet
7	170.5	30.2	94	11	09QXF2	09QXF2 mus musculi
8	77.5	13.7	92	11	09JWB5	013997 schizosacch
9	77.5	13.6	280	3	013997	013997 schizosacch
10	76	13.5	212	16	09CA622	09G622 caulobacte
11	73.5	13.0	590	10	09CAN6	09CAN6 arabidopsi
12	73	12.9	702	5	09VH96	09VH96 drosophila
13	73	12.9	1688	13	057483	057483 rana cates
14	71.5	12.7	487	10	09LVB6	09LVB6 arabidopsi
15	71.5	12.7	4848	2	007944	007944 streptomyce
16	71	12.6	303	16	09ZGB8	09ZGB8 rickettsia

17	71	12.6	530	5	09XZB3	09XZH3 amoeba prot
18	71	12.6	828	5	09YC06	09YC66 drosophila
19	71	12.6	1187	5	09V7N3	09V7N3 drosophila
20	71	12.6	1189	5	09SR60	09SR60 drosophila
21	70.5	12.5	343	5	09W212	09W212 drosophila
22	70	12.4	457	5	01B534	01B534 caenorhabdi
23	69.5	12.3	108	9	09JML7	09JML7 bacterioph
24	69.5	12.3	412	10	09M2A1	09M2A1 arabadipops
25	69.5	12.3	621	10	09ASZ8	09ASZ8 arabadipops
26	69.5	12.3	735	10	09INB5	09INB5 arabadipops
27	69.5	12.3	1048	10	09SGV7	09SGV7 arabadipops
28	69	12.2	219	17	028785	028785 archaeoglob
29	69	12.2	638	10	09W393	09W393 arabadipops
30	69	12.2	799	2	09L8Z6	09L8Z6 enterococcu
31	68.5	12.1	173	10	0816Z6	0816Z6 arabadipops
32	68.5	12.1	213	16	097M18	097M18 clostridium
33	68.5	12.1	384	5	09VCV6	09VCV6 drosophila
34	68.5	12.1	456	4	09NTG0	09NTG0 homo sapien
35	68.5	12.1	619	4	09HDC6	09HDC6 homo sapien
36	68.5	12.1	694	4	09UPP0	09UPP0 homo sapien
37	68.5	12.1	720	4	09NRR0	09NRR0 homo sapien
38	68	12.1	516	16	09A9K8	09A9K8 caulobacter
39	68	12.1	570	12	09DHH3	09DHH3 yaba-like d
40	68	12.1	4717	3	094Z48	094Z48 schizosacch
41	67	11.9	738	10	0823R0	0823R0 arabadipops
42	67	11.9	1001	15	09WC54	09WC54 human immun
43	66.5	11.8	126	5	027011	027011 tenebrio mo
44	66.5	11.8	199	13	090ZB3	090ZB3 xenopus lae
45	66.5	11.8	274	16	09PFN6	09PFN6 xyella fas

ALIGNMENTS

RESULT	1			
09GK63		PRELIMINARY;	PRT;	93 AA.
ID	09GK63			
AC	09GK63.			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	LIPOPHILIN CL2.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TaxID:9986;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LACRIMAL GLAND;			
RA	Zhao C., Nguyen T.X., Lehrer R.J.;			
RT	"Rabbit Lipophilins.";			
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF308618; AAG42806.1; -			
DR	InterPro; IPR003627; Mammaglobin_prostin.			
DR	InterPro; IPR003629; Uteroglobulin.			
DR	pfam; PF01099; uterogloblin; 1.			
DR	Prodom; PFD029354; Mammaglobin_prostin; 1.			
SQ	SEQUENCE 93 AA: 10456 MW: 4915E4SD9134AE91 CRC64;			

[illegible]


```
RESULT 2
O9GK61 PRELIMINARY: PRT: 93 AA.
AC O9GK61:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LIPOPHILIN CS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
(1)
RN SEQUENCE FROM N.A.
RP TISSUE=SUBMAXILLARY;
RC Zhao C., Nguyen T.X., Lehrer R.I.;
RT "Rabbit Lipophilins.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF308620; AAC42808.1;
DR InterPro: IPR003627; Mammaglnb_prostin.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR PRINTS: PR00486; UTEROglobln.
DR ProDom: PD029354; Mammaglnb_prostin.1.
DR SMART: SM00096; UTC; 1.
SQ SEQUENCE 93 AA; 10609 MW; E2B015AD319B2249 CRC64;
```

Query Match 41.1%; Score 232; DB 6; Length 93;
Best Local Similarity 50.0%; Pred. No. 3e-17; Indels 0; Gaps 0;
Matches 45; Conservative 20; Mismatches 25;

```
OY 14 MKLWLVMLAALLHCYADSGCKLLEDMVEKTIISDIPIPEYKELLOFIIISDAAEAMG 73
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLWLVMLAALLPLCYAGSGCPVEKMTTLLNSNSTRFYIDVKNYINDEKRELAVY 60
OY 74 KFKOCLNOSHRTLNKFGIMHTVYDSIWC 103
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 EFKNCFLSQSEETLRNVEMMETIYNSKLC 90
```

```
RESULT 3
O9GK64 PRELIMINARY: PRT: 93 AA.
AC O9GK64:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LIPOPHILIN CL.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
(1)
RN SEQUENCE FROM N.A.
RP TISSUE=LACRIMAL GLAND;
RC Zhao C., Nguyen T.X., Lehrer R.I.;
RT "Rabbit Lipophilins.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF308617; AAC42805.1;
DR InterPro: IPR003627; Mammaglnb_prostin.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR ProDom: PD029354; Mammaglnb_prostin.1.
SQ SEQUENCE 93 AA; 10350 MW; 9F9206C44372804D CRC64;
```

Query Match 40.6%; Score 229; DB 6; Length 93;
Best Local Similarity 45.6%; Pred. No. 6.2e-17; Indels 0; Gaps 0;
Matches 41; Conservative 26; Mismatches 23;

```
OY 14 MKLWLVMLAALLHCYADSGCKLLEDMVEKTIISDIPIPEYKELLOFIIISDAAEAMG 73
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 1 MKWLVVLATLPFCYAGSGCVILLESVLDKTIIDPSVSEDTTYLQKYLITDAKVALE 60
OY 74 KFKOCLNOSHRTLNKFGIMHTVYDSIWC 103
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 ELKOCFLSQSNETLANVKVAVDFDSLXC 90
RESULT 4
O9GK62 PRELIMINARY: PRT: 93 AA.
AC O9GK62:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LIPOPHILIN CP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
(1)
RN SEQUENCE FROM N.A.
RP TISSUE=PROSTATE;
RC Zhao C., Nguyen T.X., Lehrer R.I.;
RT "Rabbit Lipophilins.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF308619; AAC42807.1;
DR InterPro: IPR003627; Mammaglnb_prostin.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR ProDom: PD029354; Mammaglnb_prostin.1.
SQ SEQUENCE 93 AA; 10332 MW; C9DC35B17D372F32 CRC64;
```

Query Match 40.2%; Score 227; DB 6; Length 93;
Best Local Similarity 50.0%; Pred. No. 1e-16; Indels 0; Gaps 0;
Matches 45; Conservative 16; Mismatches 29;

```
OY 14 MKLWLVMLAALLHCYADSGCKLLEDMVEKTIISDIPIPEYKELLOFIIISDAAEAMG 73
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKCVTALMLAALLPLCYAGSGCQLDDMYTKTIDSQISLTXYINFPNNLSGMAAEAVK 60
OY 74 KFKOCLNOSHRTLNKFGIMHTVYDSIWC 103
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DKOCLFMOSNETLNINIKVFLFTVYNSPFC 90
```

```
RESULT 5
O9JHB9 PRELIMINARY: PRT: 95 AA.
AC O9JHB9:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE PROSTATIC STEROID BINDING PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
(1)
RN SEQUENCE FROM N.A.
RP MEDLINE=84057754; PubMed=6685625;
RX Hurst H.C., Parker M.G.;
RA "Rat prostatic steroid binding protein: DNA sequence and transcript  
RT maps of the two C3 genes.";
RL EMBL: V01260; CAB75892.1;
DR EMBL: V01260; CAB75892.1; JOINED.
DR EMBL: V01262; CAB75892.1; JOINED.
DR InterPro: IPR003627; Mammaglnb_prostin.
DR InterPro: IPR000329; Uteroglobln.
DR Pfam: PF01099; Uteroglobln.1.
DR ProDom: PD029354; Mammaglnb_prostin.1.
SQ SEQUENCE 95 AA; 10622 MW; B209F1BE177C52A4 CRC64;
```

```

OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LAKE VIEW; TISSUE-HARDERIAN GLAND;
RA Alvarez J., Alber J., Ashman K., Martin Alonso J.M., Dominguez P.;
RT "Cloning and differential expression of two isoforms of hetroglobin,"
RL a novel heterodimeric glycoprotein of the uteroglobin gene family.";
DR Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ252139; CAB64661.1; -.
DR InterPro: IPR003627; Mammaglobin_protstn.
DR InterPro: IPR000329; Uteroglobin.
DR Pfam: Pf01099; Uteroglobin; 1.
DR ProDom: PD029354; Mammaglobn_protstn; 1.
KW Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 94 HETEROGLOBIN B2 SUBUNIT.
FT SEQUENCE 94 AA; 10821 MW; 16C640C0674224C9 CRC64;

Query Match 30.2%; Score 170.5; DB 11; Length 94;
Best Local Similarity 37.0%; Pred. No. 1e-10;
Matches 34; Conservative 23; Mismatches 34; Indels 1; Gaps 1;

OY 14 MKLLVLMIALILHCYAD-SGCKLELDVNEKTIISDISEYKELLOEFLDSIAAAM 72
DB 1 MLVIVFPMALIPVYCRINSSGCMMDAIAATINSSVPMEEYHETVAKYTTLPYIRSTV 60
OY 73 GKFKOCFLNQSHRTLKNGMLMHVYYDSIWCN 104
DB 61 EKKECFAKQSDMDTQHNIFFVWVAVYVNSDKS 92

RESULT 8
OY1WB5 PRELIMINARY; PRT; 92 AA.
ID OY1WB5;
AC OY1WB5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE HYPOTHEICAL 10.1 KDA PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SALIVARY GLAND;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC016132; AAF16132.1; -.
KW Hypothetical protein.
SQ SEQUENCE 92 AA; 10080 MW; F1A7557E0F5568BB CRC64;

Query Match 13.7%; Score 77.5; DB 11; Length 92;
Best Local Similarity 36.4%; Pred. No. 0.75;
Matches 28; Conservative 10; Mismatches 30; Indels 9; Gaps 4;

OY 14 MKLLVLMIALILHCYADSG-CKLELDVNEKTIISDISEYKELLOEFLDSIAAAE 70
DB 1 MKLLGALLIGAAILLITSEGDGCPALQRRKVDLFLNG--TTEEVYVELKQFNENKRVLF 58
OY 71 AMGKFCFLNQSHRTL 87
DB 59 MAANIKKC---SDRTL 71

RESULT 9
ID O13997 PRELIMINARY; PRT; 280 AA.
AC O13997;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)

```


01-MAY-2000 (TrEMBLrel. 13, last sequence update)
 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 CG8358 PROTEIN.
 DE CG8358.
 GN CG8358.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richardson M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer H.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Andrews-Pfannkoch C., Baldwin D.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garb N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jallat B.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Jallat B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Mishina N.V., Mobarry C., Morris J.J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svikas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Weissman D.A., Weisslock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DB EMBL: AF003683; AAF54423.1;
 DR MEROPS: M13.0PM;
 DR FlyBase: FBgn0037727; CG8358.
 DR InterPro: IPR000718; Peptidase_M13.
 DR InterPro: IPR00130; Zn_MTPeptide.
 DR Pfam: PF01431; Peptidase_M13; 1.
 DR PRINTS: PR00766; NEPRILysin.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN.1.
 DR SEQUENCE 702 AA; 80825 MW; 4FCALAI39F1FOBB6 CRC64;

Query Match 12.9%; Score 73; DB 5; Length 702;
 Best Local Similarity 27.8%; Pred. No. 23;
 Matches 25; Conservative 19; Mismatches 20; Indels 26; Gaps 4;

DB 4 ESXKLSIMLWYVAIVLTDCEARS-----LVDRSENSDNGSSTNSRAEYO--IQ 54
 OY 8 DSSRLAKMLVLMALILHYCAADSCKLEMEVETINSDI-----SIPEYELLQ 60
 OY 61 EFIDSDAAEAAMGKFCFCFNSHRTLKNF 90
 OY 55 EY-----AEFMKSYMSQVPEPCNF 74

RESULT 13
 ID 057483 PRELIMINARY: PRT: 1688 AA.
 AC 057483;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
 DE VOLTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL, ALPHA-1S SUBUNIT
 DE (FGALPHA1S).
 OS Rana catesbeiana (Bull frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.
 NCBI_Taxid=8400;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKLETAL MUSCLE;
 RX MEDLINE=98411359; PubMed=9738021;
 RA Zhou J., Cribbs L., Yi J., Shirokov R., Perez-Reyes E., Rios E.;
 RT "Molecular cloning and functional expression of a skeletal muscle
 dihydropyridine receptor from Rana catesbeiana.";
 RL J. Biol. Chem. 273:25503-25509(1998).
 CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
 ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
 IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
 CONTRACTION, GENE EXPRESSION, CELL MOTILITY, CELL DIVISION AND
 CELL DEATH. THE ISOFORM ALPHA-1S GIVES RISE TO L-TYPE CALCIUM
 CURRENTS.
 CC -1- FUNCTION: LONG-LASTING (L-TYPE) CALCIUM CHANNELS BELONG TO THE
 "HIGH-VOLTAGE ACTIVATED (HVA)" GROUP. THEY ARE BLOCKED BY
 DIHYDROPYRIDINES (DHP), PHENYLAALKYLAMINES, BENZOTHAZEPINES, AND B
 OMEGA-AGATOXIN-IIIA (OMEGA-AGA-IIIA). THEY ARE HOWEVER INSENSITIVE
 TO OMEGA-COROTOXIN-GVIA (OMEGA-CITX-GVIA) AND OMEGA-AGATOXIN-IVA
 (OMEGA-AGA-IVA). CALCIUM CHANNELS CONTAINING THE ALPHA-1S SUBUNIT
 PLAY AN IMPORTANT ROLE IN EXCITATION-CONTRACTION COUPLING IN
 SKELETAL MUSCLE (BY SIMILARITY).
 CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
 IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE FOR-
 FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
 SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY. AN
 ADDITIONAL GAMMA SUBUNIT IS PRESENT ONLY IN SKELETAL MUSCLE L-TYPE
 CHANNEL (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE SPECIFIC.
 CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
 HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
 POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
 PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
 SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
 CC -1- DOMAIN: THE LOOP BETWEEN REPEATS II AND III INTERACTS WITH THE
 RYANODINE RECEPTOR, AND IS THEREFORE IMPORTANT FOR CALCIUM RELEASE
 FROM THE ENDOPLASMIC RETICULUM NECESSARY FOR MUSCLE CONTRACTION
 (BY SIMILARITY).
 CC -1- PTM: PHOSPHORYLATION BY CAPK STIMULATES THE CALCIUM CHANNEL
 FUNCTION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
 FAMILY.
 DR EMBL: AF037625; AAC36126.1;
 DR InterPro: IPR000636; Cation_chan_nou_11g.
 DR InterPro: IPR002111; Cal_channel_TripL.
 DR InterPro: IPR002077; Ca_channel.
 DR InterPro: IPR001682; Channel_pore_Ca_Na.
 DR InterPro: IPR001064; Crystal11in.
 DR Pfam: PF00520; ion_trans_4.
 DR PRINTS: PR00167; CACCHANNEL.
 DR PROSITE: PS00225; CRISTALLIN_BETAGAMMA; UNKNOWN.1.
 DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 KW Calcium channel; Glycoprotein; Repeat; Multigene family;
 KW Calcium-binding; Phosphorylation.

FT	REPEAT	38	335	I.
FT	REPEAT	417	663	II.
FT	REPEAT	784	1066	III.
FT	REPEAT	1103	1371	IV.
FT	DOMAIN	1	51	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	52	70	S1 OF REPEAT I (POTENTIAL).
FT	DOMAIN	71	88	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	89	108	S2 OF REPEAT I (POTENTIAL).
FT	DOMAIN	109	120	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	121	139	S3 OF REPEAT I (POTENTIAL).
FT	DOMAIN	140	158	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	159	177	S4 OF REPEAT I (POTENTIAL).
FT	DOMAIN	178	196	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	197	216	S5 OF REPEAT I (POTENTIAL).
FT	DOMAIN	217	307	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	308	332	S6 OF REPEAT I (POTENTIAL).
FT	DOMAIN	333	431	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	432	450	S1 OF REPEAT II (POTENTIAL).
FT	DOMAIN	451	465	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	466	485	S2 OF REPEAT II (POTENTIAL).
FT	DOMAIN	486	493	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	494	512	S3 OF REPEAT II (POTENTIAL).
FT	DOMAIN	513	522	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	523	541	S4 OF REPEAT II (POTENTIAL).
FT	DOMAIN	542	560	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	561	580	S5 OF REPEAT II (POTENTIAL).
FT	DOMAIN	581	635	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	636	660	S6 OF REPEAT II (POTENTIAL).
FT	DOMAIN	661	797	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	798	816	S1 OF REPEAT III (POTENTIAL).
FT	DOMAIN	817	832	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	833	852	S2 OF REPEAT III (POTENTIAL).
FT	DOMAIN	853	864	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	865	883	S3 OF REPEAT III (POTENTIAL).
FT	DOMAIN	884	909	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	891	909	S4 OF REPEAT III (POTENTIAL).
FT	DOMAIN	910	928	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	929	948	S5 OF REPEAT III (POTENTIAL).
FT	DOMAIN	949	1038	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1039	1063	S6 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1064	1116	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1117	1135	S1 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1136	1150	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1151	1170	S2 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1171	1178	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1179	1197	S3 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1198	1218	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1219	1237	S4 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1238	1256	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1257	1276	S5 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1277	1343	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1344	1368	S6 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1369	1688	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	22	28	POLY-PRO.
FT	DOMAIN	438	441	POLY-TLE.
FT	DOMAIN	561	567	POLY-LEU.
FT	DOMAIN	744	749	POLY-GLU.
FT	DOMAIN	1552	1559	POLY-GLU.
FT	DOMAIN	355	372	BINDING TO THE BETA SUBUNIT (BY SIMILARITY).
FT	SITE	290	290	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	SITE	612	612	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	SITE	1012	1012	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	SITE	1310	1310	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	SITE	1310	1310	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	BINDING	986	1075	TO DIHYDROPYRIDINES (BY SIMILARITY).
FT	BINDING	1324	1390	TO DIHYDROPYRIDINES (BY SIMILARITY).
FT	BINDING	1336	1379	TO PHENALKYLAMINES (BY SIMILARITY).
FT	MOD_BIND	1379	1379	PHOSPHORYLATION (BY CAPR) (POTENTIAL).
FT	CA_BIND	1397	1408	BY SIMILARITY.

```

FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
TT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1688 AA; 192421 MW; 58BA05B36823017A CRC64;

Query Match 12.9%; Score 73; DB 13; Length 1688;
Best Local Similarity 24.1%; Pred. No. 64;
Matches 28; Conservative 16; Mismatches 36; Indels 36; Gaps 5;

OY 16 LLMVLMIAALLHCACDAGCKLDEWVKRTI--NSDISIPEYKELLQEIFDS--AAAE 70
DB 194 LHLHLLLEMMITIAIVGLLEFSKMKRTCYCFKDITITAT-----VDNEKRPACS 244
OY 71 AMGKRCQCFLNOSH-----RTKNGGLMHTVDSI-----WCN 104
DB 245 SFGCGROCSINSECRGWWPGPNNGITFHFDNKGFMALTVYOCIMTEGTEVLVWVN 300

RESULT 14
O9LV86 PRELIMINARY; PRT; 487 AA.
AC O9LV86;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 5, P1 CLONE:HXK3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S.,
RT Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones. *;
RL DNA Res. 7:31-63(2000).
DR EMBL: AB019236; BAA97303.1;
SQ SEQUENCE 487 AA; 54862 MW; 632BAVE298807D27 CRC64;

Query Match 12.7%; Score 71.5; DB 10; Length 487;
Best Local Similarity 22.7%; Pred. No. 22;
Matches 17; Conservative 16; Mismatches 33; Indels 9; Gaps 1;

OY 27 LHCYADSGCKLDEWVEKTI NSDISIPEYKELLQEIFDS-----DAAAEAMGKFRQ 77
DB 366 VCEQYDMWKSVEIEFDNSLINEVLLPEQVAFEEFEVKEKSDKAMAAYDTAQDAIEKARE 425
OY 78 CFLNDSHRTIAKNGL 92
DB 426 GLSEETKKAQEMRL 440

RESULT 15
O07944 PRELIMINARY; PRT; 4848 AA.
AC O07944;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PRISTINAMYCIN I SYNTHASE 3 AND 4.
DE SBNDE.
OS Streptomyces pristinaespiralis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=38300;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-SP92;
 RA de Crecy-Lagard V.A., Saurin W., Thibaut D., Gil P., Naudin L.,
 RA Crouzet J., Blanc V.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SP92;
 RX MEDLINE-97158664; Pubmed-9006024;
 RA De Crecy-Lagard V., Blanc V., Gil P., Naudin L., Lorenzon S.,
 RA Farnetion A., Bamas-Jacques N., Crouzet J., Thibaut D.;
 RT *Pristinamycin I biosynthesis in Streptomyces pristinaespiralis:
 RT molecular characterization of the first two structural peptide
 RT synthetase genes.*;
 RL J. Bacteriol. 179:705-713(1997).
 DR EMBL: Y11548; CAA72312.1; -;
 DR EMBL: X98690; CAA67249.1; -;
 DR HSSP; P14687; 1AMU
 DR InterPro: IPR002106; AA_LRNALigase_II.
 DR InterPro: IPR000873; AMP-bind.
 DR Interf 3; IPR001242; DUF4.
 DR InterPro: IPR002336; Erythecurin.
 DR InterPro: IPR000379; Est_1ip_thioest_actsite.
 DR InterPro: IPR001601; Meth-transf.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR InterPro: IPR000051; SAM_bind.
 DR InterPro: IPR01031; Thioesterase.
 DR Pfam: PF00501; AMP-binding; 4.
 DR Pfam: PF00668; Condensation; 5.
 DR Pfam: PF00550; PP-binding; 4.
 DR Pfam: PF00975; Thioesterase; 1.
 DR PRINTS; PRO0154; AMPBINDING.
 DR PRINTS; PRO0611; ERYTHECURIN.
 DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
 DR PROSITE; PS50075; ACP_DOMAIN; 4.
 DR PROSITE; PS00455; AMP_BINDING; 4.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
 KM Phosphopantetheine.
 SQ SEQUENCE 4848 AA; 522081 MW; E17591617A2B9A0E CRC64;

Query Match 12.7%; Score 71.5; DB 2; Length 4848;
 Best Local Similarity 32.6%; Pred.No. 3.1e+02;
 Matches 29; Conservative 10; Mismatches 31; Indels 19; Gaps 3;
 QY 2 HARLMTDSSRLAMKLLMYLM--LAALLHGYADSCCKL-----IEDMVEKPTINS 48
 DB 2716 HORLMAATGCRASLFMWLOAAFAALTRHGAGTIDPLGSPACRTDADLEDVVGFFVNT 2775
 QY 49 DI-----SIPYKELIOEFIDSDAAAE 71
 DB 2776 LVKRTDTSGDPTFRELLQVRRESDLAAYN 2804

Search completed: June 20, 2002, 11:05:59
 Job time: 230 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 11:01:45 ; Search time 17.99 Seconds
(without alignments)
526.499 Million cell updates/sec

Title: US-09-673-395A-238
Perfect score: 564
Sequence: 1 CHARLNTDSSRLMKLMLV.....NFGLMHTVDSIMCNKSN 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 265615 seqs, 87701152 residues

Total number of hits satisfying chosen parameters: 265615

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending-Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/US06_NEW.COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US07_NEW.COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US08_NEW.COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US09_NEW.COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US10_NEW.COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US50_NEW.COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US50_NEW.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	496	87.9	95	1	PCT-US02-07826-187
2	496	87.9	95	6	US-10-097-340-187
3	283	50.2	93	6	US-10-119-480-58
4	281	49.8	93	1	PCT-US02-12378-503
5	281	49.8	93	6	US-10-124-805-503
6	281	49.8	410	1	PCT-US02-12378-495
7	281	49.8	410	6	US-10-124-805-495
8	281	49.8	743	1	PCT-US02-12378-494
9	281	49.8	743	6	US-10-124-805-494
10	281	49.8	1095	1	PCT-US02-12378-493
11	281	49.8	1095	6	US-10-124-805-493
12	206	36.5	220	5	US-09-684-215A-10
13	77	13.7	21	1	PCT-US02-12378-498
14	77	13.7	21	6	US-10-124-805-498
15	70	12.4	457	7	US-60-360-039-6769
16	68	12.1	1071	7	US-60-360-039-22283
17	67.5	12.0	871	7	US-60-360-039-22330
18	66	11.7	20	1	PCT-US02-12378-489
19	66	11.7	20	6	US-10-124-805-499
20	66	11.7	135	5	US-09-540-209B-8278
21	66	11.7	293	5	US-09-902-525-42
22	65.5	11.6	429	5	US-09-573-655B-1310
23	64.5	11.4	395	5	US-09-935-625-7865
24	64.5	11.4	446	5	US-09-935-625-7865
25	64.5	11.4	610	5	US-09-935-625-7864
26	64	11.3	1280	5	US-09-935-625-29952

27	64	11.3	1310	5	US-09-935-625-29951	Sequence 29951, A
28	64	11.3	1389	5	US-09-935-625-29950	Sequence 29950, A
29	63.5	11.3	246	6	US-10-104-047-3660	Sequence 3660, Ap
30	63.5	11.3	507	6	US-09-540-209B-8498	Sequence 8498, Ap
31	63.5	11.3	1360	5	US-09-573-655B-282	Sequence 282, App
32	63	11.2	410	7	US-60-360-039-10896	Sequence 10896, A
33	62.5	11.1	1094	5	US-09-573-655B-1388	Sequence 1388, A
34	62	11.0	193	5	US-09-540-209B-10334	Sequence 10334, A
35	62	11.0	193	5	US-09-540-209B-10352	Sequence 10352, A
36	62	11.0	472	5	US-09-540-209B-5684	Sequence 5684, Ap
37	62	11.0	569	5	US-09-573-655B-216	Sequence 216, App
38	62	11.0	1204	5	US-09-935-625-30799	Sequence 30799, A
39	62	11.0	1219	5	US-09-935-625-30802	Sequence 30802, A
40	62	11.0	1234	5	US-09-935-625-30798	Sequence 30798, A
41	62	11.0	1313	5	US-09-935-625-30797	Sequence 30797, A
42	62	11.0	1384	5	US-09-935-625-30801	Sequence 30801, A
43	61.5	10.9	422	5	US-09-573-655B-1311	Sequence 1311, Ap
44	61.5	10.9	734	7	US-60-360-039-3017	Sequence 3017, Ap
45	61.5	10.9	915	1	PCT-US02-13142-3197	Sequence 3197, Ap

ALIGNMENTS

RESULT 1
PCT-US02-07826-187
Sequence 187, Application PC/TUS0207826
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification, File Reference: MRI-0303C
FILE REFERENCE: MRI-0303C
CURRENT APPLICATION NUMBER: PCT/US02/07826
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 187
LENGTH: 95
TYPE: PCT
ORGANISM: Homo sapiens
PCT-US02-07826-187

Query Match 87.9%; Score 496; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 7.7e-47;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 14 MKLWVLMALALLHCYADSGCKLLEDAVEKTIINDISIPKELLOFIDDAAEAMG 73
DB 1 MKLWVLMALALLHCYADSGCKLLEDAVEKTIINDISIPKELLOFIDDAAEAMG 60
DB 74 KFKOCFLNOSHRTLNKFGMLMHTVDSIMCNKSN 108
DB 61 KFKOCFLNOSHRTLNKFGMLMHTVDSIMCNKSN 95

RESULT 2
US-10-097-340-187
Sequence 187, Application US/10097340

```

; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVARAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, JR.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xunel ZHAO
; APPLICANT: Karen GLANT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins for the Identification,
; FILE REFERENCE: MRI-030
; FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
; CURRENT APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
; HS-10-097-340-187

Query Match      87.9%; Score 496; DB 6; Length 95;
Best Local Similarity 100.0%; Pred. No. 7.7e-47;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLWVLMALLLHGYADSGCKLLEDVVEKTIINSIPYKELLOEFIDSDAAAEAMG 73
DB 1 MKLWVLMALLLHGYADSGCKLLEDVVEKTIINSIPYKELLOEFIDSDAAAEAMG 60

QY 74 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNKN 108
DB 61 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNKN 95

RESULT 3
US-10-119-480-58
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
```

```

; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3530P1C1
; CURRENT APPLICATION NUMBER: US/10/119,480
; CURRENT FILING DATE: 2002-04-09
; NUMBER OF SEQ ID NOS: 246
; Prior Application removed - See file Wrapper or Palm
; SEQ ID NO 58
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-119-480-58

Query Match      50.2%; Score 283; DB 6; Length 93;
Best Local Similarity 57.8%; Pred. No. 1.1e-23;
Matches 52; Conservative 20; Mismatches 18; Indels 0; Gaps 0;

QY 14 MKLWVLMALLLHGYADSGCKLLEDVVEKTIINSIPYKELLOEFIDSDAAAEAMG 73
DB 1 MKVYVLLALALPLXYAGSGCVLLESVEKTIIDPSVSEYKADLQRFIDTEQEAIVE 60

QY 74 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWC 103
DB 61 EFKECFLSOSNETLANFRVWHTIYDSLXC 90

RESULT 4
PCT-US02-12378-503
; Sequence 503, Application PC/TUS0212378
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margareta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.47003PC
; CURRENT APPLICATION NUMBER: PCT/US02/12378
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 503
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-12378-503

Query Match      49.8%; Score 281; DB 1; Length 93;
Best Local Similarity 58.7%; Pred. No. 1.8e-23;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 14 MKLWVLMALLLHGYADSGCKLLEDVVEKTIINSIPYKELLOEFIDSDAAAEAMG 73
DB 1 MKLWVLMALALSHQHCYAGSGCPLLENVISTKINPOVSKTEYKELLOEFIDNATTNAID 60

QY 74 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCN 105
DB 61 ELKBCFLNQTDETLNVEFMQLIYDSLCLD 92

RESULT 5
US-10-124-805-503
```



```

; TITLE OF INVENTION: Methods of Using a Mycobacterium tuberculosis Coding
; TITLE OF INVENTION: Sequence to Facilitate Stable and High Yield Expression
; GENERAL INFORMATION: of Heterologous Proteins
; FILE REFERENCE: 014058-008010US
; CURRENT APPLICATION NUMBER: US/09/684,215A
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/158,585
; PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra12-human mamaglobin
; OTHER INFORMATION: fusion polypeptide
; US-09-684-215/-10
```

```

Query Match          36.5%; Score 206; DB 5; Length 220;
Best Local Similarity 52.1%; Pred. No. 7e-15;
Matches 38; Conservative 13; Mismatches 22; Indels 0; Gaps 0;
```

```

QY 33 SGCKLEDMYKRTINSIPETKELQERIDSDAAEAMGKFKOCLNOSHRTLNFG 92
    ||| |||:: |||:: | ||| ||| |||:: |:: |:: |||:: || |::
Db 147 SGCPLENVISKRTIPQVSKTEYKELQERIDNATNNAIDELKECPINOTDETLNVEV 206
```

```

QY 93 MMHTVYDSIMCNM 105
    |:: ||| ||::
Db 207 FMQILYDSLCDL 219
```

```

RESULT 13
PCT-US02-12378-498
; Sequence 498, Application PC/TUS0212378
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.47003PC
; CURRENT APPLICATION NUMBER: PCT/US02/12378
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 498
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-12378-498
```

```

Query Match          13.7%; Score 77; DB 1; Length 21;
Best Local Similarity 68.4%; Pred. No. 0.045;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 28 HCYADSGCKLHDMYKRTI 46
    ||| ||| ||| |||:: |||
Db 3 HCYAGSGCPLENVISKTI 21
```

```

RESULT 14
US-10-124-805-498
; Sequence 498, Application US/10124805
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.4700312
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 498
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-124-805-498
```

```

Query Match          13.7%; Score 77; DB 6; Length 21;
Best Local Similarity 68.4%; Pred. No. 0.045;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 28 HCYADSGCKLHDMYKRTI 46
    ||| ||| ||| |||:: |||
Db 3 HCYAGSGCPLENVISKTI 21
```

```

RESULT 15
US-60-360-039-6769
; Sequence 6769, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6769
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-60-360-039-6769
```

```

Query Match          12.4%; Score 70; DB 7; Length 457;
Best Local Similarity 28.3%; Pred. No. 10;
Matches 30; Conservative 16; Mismatches 32; Indels 28; Gaps 6;
```

```

QY 1 CHARLNTDSSRLAMKILMVLMLA-----ALLHCVADSGCKL-----LEDMYKRT-- 45
    |:: |:: ||| |:: ||| |:: ||| |:: ||| |:: ||| |:: |||
Db 178 CTEKMEKSKSRRA--YNYMYLAFQVYFVAFVMAFCYANITYSVLSKRAQRTIRRMVETSA 235
    |:: |:: ||| |:: ||| |:: ||| |:: ||| |:: ||| |:: |||
QY 46 INSDISIP-----EYKELQERIDSDAAEAMGKFKOCLNOSHRT 86
    |:: |:: ||| |:: ||| |:: ||| |:: ||| |:: ||| |:: |||
Db 236 LESSCAFPVHGLEOYENELNEFLDKQ-----KEKORVVLQNRRT 275
    |:: |:: ||| |:: ||| |:: ||| |:: ||| |:: ||| |:: |||
```

```

Search completed: June 20, 2002, 11:03:31
Job time: 106 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 11:01:44 ; Search time 107.96 Seconds

(Without alignments)
352.111 Million cell updates/sec

Title: US-09-673-395A-238

Perfect score: 564
Sequence: 1 CHARLNTDSSRLAMKLIWLV.....NEGLMHHTVYDSIWMKSN 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US103_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US104_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US105_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US106_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US107_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US108_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US109_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US110_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US111_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	564	100.0	108	US-09-673-395A-238
2	496	87.9	95	PCT-US01-09339-3
3	496	87.9	95	PCT-US96-03857-6
4	496	87.9	95	PCT-US99-22753-2
5	496	87.9	95	US-08-951-750-1
6	496	87.9	95	US-09-110-716-31
7	496	87.9	95	US-09-412-231-12

8	496	87.9	95	US-09-471-276-1555	Sequence 1555, Ap
9	496	87.9	95	US-09-549-342A-3	Sequence 3, Appli
10	496	87.9	95	US-09-985-911-6	Sequence 6, Appli
11	475	84.2	92	US-09-057-719-1075	Sequence 1075, Ap
12	408	72.3	77	US-09-110-716-13	Sequence 13, Appli
13	393.5	69.8	76	US-09-110-716-40	Sequence 40, Appli
14	390	69.1	74	US-09-367-009-3	Sequence 3, Appli
15	343	60.8	71	US-60-234-446-797	Sequence 797, App
16	318	56.4	62	US-60-160-203-3646	Sequence 3646, Ap
17	281	49.8	93	PCT-US01-04439-1	Sequence 1, Appli
18	281	49.8	93	PCT-US01-04439-34	Sequence 34, Appli
19	281	49.8	93	PCT-US98-17191-2	Sequence 2, Appli
20	281	49.8	93	PCT-US99-22616-2	Sequence 2, Appli
21	281	49.8	93	US-08-697-106-2	Sequence 2, Appli
22	281	49.8	93	US-08-722-304-2	Sequence 2, Appli
23	281	49.8	93	US-08-747-547-3	Sequence 3, Appli
24	281	49.8	93	US-08-747-547-10	Sequence 10, Appli
25	281	49.8	93	US-08-912-149-17	Sequence 17, Appli
26	281	49.8	93	US-08-951-750-3	Sequence 3, Appli
27	281	49.8	93	US-09-162-622-2	Sequence 2, Appli
28	281	49.8	93	US-09-215-818-5	Sequence 5, Appli
29	281	49.8	93	US-09-467-602-5	Sequence 5, Appli
30	281	49.8	93	US-09-467-602A-5	Sequence 5, Appli
31	281	49.8	93	US-09-509-015-2	Sequence 2, Appli
32	281	49.8	93	US-09-534-292-17	Sequence 17, Appli
33	281	49.8	93	US-09-580-376-27	Sequence 27, Appli
34	281	49.8	93	US-09-757-417-27	Sequence 27, Appli
35	281	49.8	93	US-09-780-842-1	Sequence 1, Appli
36	281	49.8	93	US-09-780-842-34	Sequence 34, Appli
37	281	49.8	93	US-09-834-759-503	Sequence 503, App
38	281	49.8	93	US-09-905-673-1	Sequence 1, Appli
39	281	49.8	93	US-09-905-673-34	Sequence 34, Appli
40	281	49.8	93	US-09-934-054-3	Sequence 3, Appli
41	281	49.8	93	US-09-934-054-10	Sequence 10, Appli
42	281	49.8	93	US-10-007-805-503	Sequence 503, App
43	281	49.8	93	US-10-042-945-27	Sequence 27, Appli
44	281	49.8	93	US-10-076-622-503	Sequence 503, App
45	281	49.8	93	US-10-096-319-1	Sequence 1, Appli

ALIGNMENTS

US-09-673-395A-238
Sequence 238, Application US/09673395A
GENERAL INFORMATION:
APPLICANT: SPECHT, THOMAS
APPLICANT: HINZMAN, BERNI
APPLICANT: SCHMITT, ARMIN
APPLICANT: PILASKI, CHRISTIAN
APPLICANT: DAHL, EDGAR
TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
FILE REFERENCE: ALBRE-12
CURRENT APPLICATION NUMBER: US/09/673, 395A
CURRENT FILING DATE: 2000-10-17
NUMBER OF SEQ ID NOS: 637
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 238
LENGTH: 108
TYPE: PRT
ORGANISM: Homo sapiens
US-09-673-395A-238

Query Match 100.0%; Score 564; D: 20; Length 108;
Best Local Similarity 100.0%; Pred. No. 5.2e+57;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CHARLNTDSSRLAMKLIWLVLAALLHCHYADSGCKLEDMVEKTIINSISIFREYELLO 60
DB 1 CHARLNTDSSRLAMKLIWLVLAALLHCHYADSGCKLEDMVEKTIINSISIFREYELLO 60

QY	61 EPTDSDAAAEANGKFKCOFLNOSHHTLNFGLMMHVTYDSIWCNNKSN 108
Db	61 EPTDSDAAAEANGKFKCOFLNOSHHTLNFGLMMHVTYDSIWCNNKSN 108

RESULT 2
PCT-US01-09339-3
; Sequence 3, Application PC/TUS0109339

```

; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 3
;
; LENGTH: 95
;
; TYPE: PRF
;
; ORGANISM: Homo sapiens
;
PCF-US01-09339-3

```

[illegible]

RESULT 3
PCT-US96-03857-6
: Sequence 6, Application PC/TUS9603857
: GENERAL INFORMATION:
: APPLICANT: Gentz, Reiner
: TITLE OF INVENTION: HUMAN ENDOMETRIAL SPECIFIC
: TITLE OF INVENTION: STEROID-BINDING FACTOR I, II AND III
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
: ADDRESSEE: STEWART & OLSTEIN
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068-1739
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: PCT/US96/03857
3 FILING DATE:
4 CLASSIFICATION:
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Ferraro, Gregory D
7 REGISTRATION NUMBER: 56,134
8 REFERENCE/DOCKET NUMBER: 325800-5200
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: 201-994-1744
11 TELEFAX: 201-994-1740
12 INFORMATION FOR SEQ ID NO: 6:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 95 amino acids
15 TYPE: amino acid
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18
19 PCT-US96-03857-6

```

[illegible]

```

: RESULT 4
: PCT-US99-22753-2
: Sequence 2, Application PC/US9922753
: GENERAL INFORMATION:
: APPLICANT: Macina, Roberto A.
: APPLICANT: DIADEXUS LLC
: TITLE OF INVENTION: A Novel Method of Diagnosing, Monitoring, Staging
: TITLE OF INVENTION: Imaging and Treating Gynecologic Cancers
: FILE REFERENCE: DEX-0045
: CURRENT APPLICATION NUMBER: PCT/US99/22753
: CURRENT FILING DATE: 1999-09-30
: EARLIER APPLICATION NUMBER: 60/102,743
: EARLIER FILING DATE: 1998-10-02
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 2
: LENGTH: 95
: TYPE: PRT
: ORGANISM: Homo sapiens
: PCT-US99-22753-2

```

[illegible]

RESULT 5
US-08-951-750-1
; Sequence 1, Application US/08951750
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

RESULT 5
US-08-951-750-1
; Sequence 1, Application US/08951750
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

```

; APPLICANT: Shah, Purvi
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: NOVEL MAMMOGLOBIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Inocyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; City: Palo Alto,
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,750
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0381 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTN0105
; CLONE: 2295453
;
US-08-951-750-1

Query Match      87.9%; Score 496; DB 13; Length 95;
Best Local Similarity 100.0%; Pred. No. 3,4e-49;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLWVLMALALLHCYADSGCKLEDMVEKTI NSDISIPYKELLOEFITSDAAAEAMG 73
      |||||||
DB 1 MKLWVLMALALLHCYADSGCKLEDMVEKTI NSDISIPYKELLOEFITSDAAAEAMG 60

QY 74 KFKCFLNOSHRTLNKNGIMMHTVYDSTWCMNKS 108
      |||||||
DB 61 KFKCFLNOSHRTLNKNGIMMHTVYDSTWCMNKS 95

RESULT 6
US-09-110-716-31
; Sequence 31, Application US/09110716A
; GENERAL INFORMATION:
; APPLICANT: Lehrer, Robert I.
; APPLICANT: Zhao, Chengquan
; APPLICANT: Glasgow, Benjamin J.
; TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
; FILE REFERENCE: 22000-20596 00
; CURRENT APPLICATION NUMBER: US/09/110,716A
; EARLIER FILING DATE: 1998-07-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Hapophyllin C
;
US-09-110-716-31
```

```

Query Match      87.9%; Score 496; DB 15; Length 95;
Best Local Similarity 100.0%; Pred. No. 3,4e-49;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLWVLMALALLHCYADSGCKLEDMVEKTI NSDISIPYKELLOEFITSDAAAEAMG 73
      |||||||
DB 1 MKLWVLMALALLHCYADSGCKLEDMVEKTI NSDISIPYKELLOEFITSDAAAEAMG 60

QY 74 KFKCFLNOSHRTLNKNGIMMHTVYDSTWCMNKS 108
      |||||||
DB 61 KFKCFLNOSHRTLNKNGIMMHTVYDSTWCMNKS 95

RESULT 7
US-09-412-231-12
; Sequence 12, Application US/09412231
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: NOVEL SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING
; FILE REFERENCE: CURA-29 secreted proteins
; CURRENT APPLICATION NUMBER: US/09/412,231
; EARLIER FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: USSN 60/103,195
; PRIOR FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-412-231-12

Query Match      87.9%; Score 496; DB 18; Length 95;
Best Local Similarity 100.0%; Pred. No. 3,4e-49;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLWVLMALALLHCYADSGCKLEDMVEKTI NSDISIPYKELLOEFITSDAAAEAMG 73
      |||||||
DB 1 MKLWVLMALALLHCYADSGCKLEDMVEKTI NSDISIPYKELLOEFITSDAAAEAMG 60

QY 74 KFKCFLNOSHRTLNKNGIMMHTVYDSTWCMNKS 108
      |||||||
DB 61 KFKCFLNOSHRTLNKNGIMMHTVYDSTWCMNKS 95

RESULT 8
US-09-471-276-1555
; Sequence 1555, Application US/09471276
; GENERAL INFORMATION:
; APPLICANT: Dumes Mline Edwards, J.B.
; APPLICANT: Duclert A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: GENSET.025CP1
; CURRENT APPLICATION NUMBER: US/09/471,276
; EARLIER FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1622
; SOFTWARE: Patent. pm
; SEQ ID NO 1555
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL.
```

LOCATION: -18.-1
US-09-471-276-1555

Query Match 87.9%; Score 496; DB 18; Length 95;
Best Local Similarity 100.0%; Pred. No. 3,4e-49;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLWVLMALALHCHYADSGCKLLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 73
DB 1 MKLWVLMALALHCHYADSGCKLLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 60

QY 74 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNMKSN 108
DB 61 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNMKSN 95

RESULT 9
US-09-549-342A-3
Sequence 3, Application US/09549342A
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Colpitts, Tracey L.
APPLICANT: Russell, John C.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR
FILE REFERENCE: 5972.US.P6
CURRENT APPLICATION NUMBER: US/09/549,342A
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: US 09/467,602
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: US 09/215,818
PRIOR FILING DATE: 1998-12-18
PRIOR APPLICATION NUMBER: US 08/912,276
PRIOR FILING DATE: 1997-08-17
PRIOR APPLICATION NUMBER: US 08/697,105
PRIOR FILING DATE: 1996-08-19
PRIOR APPLICATION NUMBER: US 08/912,149
PRIOR FILING DATE: 1997-08-15
PRIOR APPLICATION NUMBER: US 08/697,106
PRIOR FILING DATE: 1996-08-19
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 95
TYPE: PRT
ORGANISM: Homo sapiens
US-09-549-342A-3

Query Match 87.9%; Score 496; DB 19; Length 95;
Best Local Similarity 100.0%; Pred. No. 3,4e-49;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLWVLMALALHCHYADSGCKLLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 73
DB 1 MKLWVLMALALHCHYADSGCKLLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 60

QY 74 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNMKSN 108
DB 61 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNMKSN 95

RESULT 10
US-09-985-911-6
Sequence 6, Application US/09985911
GENERAL INFORMATION:
APPLICANT: NI ET AL.
TITLE OF INVENTION: HUMAN ENDOMETRIAL SPECIFIC STEROID-BINDING FACTOR I, II AND III
FILE REFERENCE: PF257D3
CURRENT APPLICATION NUMBER: US/09/985,911
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: 09/583,169

PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 09/263,810
PRIOR FILING DATE: 1999-03-06
PRIOR APPLICATION NUMBER: 08/821,451
PRIOR FILING DATE: 1997-03-21
PRIOR APPLICATION NUMBER: 60/014,724
PRIOR FILING DATE: 1996-03-21
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 95
TYPE: PRT
ORGANISM: human
US-09-985-911-6

Query Match 87.9%; Score 496; DB 23; Length 95;
Best Local Similarity 100.0%; Pred. No. 3,4e-49;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLWVLMALALHCHYADSGCKLLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 73
DB 1 MKLWVLMALALHCHYADSGCKLLEDVMEKTIINSISIPYKELLOEFIDSDAAAEAMG 60

QY 74 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNMKSN 108
DB 61 KFKOCFLNOSHRTLKNFGMLMHTVYDSIWCNMKSN 95

RESULT 11
US-09-057-719-1075
Sequence 1075, Application US/09057719
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste Dumas Milne
APPLICANT: Duclet, Aymeric
TITLE OF INVENTION: EXPRESSED SEQUENCE TAGS FOR SECRETED
NUMBER OF SEQUENCES: 1207
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/057,719
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: GENSET 025A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1075:
SEQUENCE CHARACTERISTICS:
LENGTH: 92 amino acids
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Cancerous prostate
FEATURE:
NAME/KEY: sig_peptide
LOCATION: -18.-1

IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 10.3
OTHER INFORMATION: seq VMLTALLHCYA/DS
US-09-057-719-1075

Query Match 84.2%; Score 475; DB 14; Length 92;
Best Local Similarity 98.9%; Pred. No. 9e-47;
Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 14 MKLMTVMTLALLHCYADSCCKLEDMVEKTI NSDISIPEYKELLOEFTIDSDAAAEAMG 73
DB 1 MKLMTVMTLALLHCYADSCCKLEDMVXKTI NSDISIPEYKELLOEFTIDSDAAAEAMG 60

OY 74 KFKOCFLNOSHRTLKNGFLMHTVYDSIWCNM 105
DB 61 KFKOCFLNOSHRTLKNGFLMHTVYDSIWCNM 92

RESULT 12
US-09-110-716-13
Sequence 13, Application US/09110716A
GENERAL INFORMATION:

APPLICANT: Lehrer, Robert I.
APPLICANT: Zhao, Chengquan
APPLICANT: Glasgow, Benjamin J.
TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
FILE REFERENCE: 22000-20596.00
CURRENT APPLICATION NUMBER: US/09/110.716A
CURRENT FILING DATE: 1998-07-07
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 77
TYPE: PRT
ORGANISM: lipophilin C
US-09-110-716-13

Query Match 72.3%; Score 408; DB 15; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.3e-39;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 DSCCKLEDMVEKTI NSDISIPEYKELLOEFTIDSDAAAEAMGKFKOCFLNOSHRTLKNG 91
DB 1 DSCCKLEDMVEKTI NSDISIPEYKELLOEFTIDSDAAAEAMGKFKOCFLNOSHRTLKNG 60

OY 92 LMHTVYDSIWCNMKN 108
DB 61 LMHTVYDSIWCNMKN 77

RESULT 13
US-09-110-716-40
Sequence 40, Application US/09110716A
GENERAL INFORMATION:

APPLICANT: Lehrer, Robert I.
APPLICANT: Zhao, Chengquan
APPLICANT: Glasgow, Benjamin J.
TITLE OF INVENTION: PEPTIDES CHARACTERISTIC OF CERTAIN TUMORS
FILE REFERENCE: 22000-20596.00
CURRENT APPLICATION NUMBER: US/09/110.716A
CURRENT FILING DATE: 1998-07-07
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 40
LENGTH: 76
TYPE: PRT
ORGANISM: Lpnc
US-09-110-716-40

Query Match 69.8%; Score 393.5; DB 15; Length 76;

Best Local Similarity 98.7%; Pred. No. 2e-37;
Matches 76; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 32 DSCCKLEDMVEKTI NSDISIPEYKELLOEFTIDSDAAAEAMGKFKOCFLNOSHRTLKNG 91
DB 1 DSCCKLEDMVEKTI NSDISIPEYKELLOEFTIDSDAAAEAMGKFKOCFLNOSHRTLKNG 60

OY 92 LMHTVYDSIWCNMKN 108
DB 61 LMHTVYDSIWCNMKN 76

RESULT 14
US-09-367-009-3
Sequence 3, Application US/09367009
GENERAL INFORMATION:

APPLICANT: Morris, Carol
APPLICANT: Wilcox, Mark
APPLICANT: Bolis, Shirley
APPLICANT: Walsh, Bradley
APPLICANT: Herbert, Ben
APPLICANT: Molloy, Mark
APPLICANT: Gooley, Andrew Arthur
APPLICANT: Williams, Keith Leslie
TITLE OF INVENTION: Diagnosis of Disease Using Tears
FILE REFERENCE: 47763-5010-US
CURRENT APPLICATION NUMBER: US/09/367.009
CURRENT FILING DATE: 1999-11-08
PRIOR APPLICATION NUMBER: AU P05009
PRIOR FILING DATE: 1997-02-07
PRIOR APPLICATION NUMBER: PCT/AU98/00071
PRIOR FILING DATE: 1998-02-06
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 74
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Disease marker sequence
US-09-367-009-3

Query Match 69.1%; Score 390; DB 17; Length 74;
Best Local Similarity 98.6%; Pred. No. 5e-37;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 32 DSCCKLEDMVEKTI NSDISIPEYKELLOEFTIDSDAAAEAMGKFKOCFLNOSHRTLKNG 91
DB 1 DSCCKLEDMVEKTI NSDISIPEYKELLOEFTIDSDAAAEAMGKFKOCFLNOSHRTLKNG 60

OY 92 LMHTVYDSIWCNM 105
DB 61 LMHTVYDSIWCNM 74

RESULT 15
US-60-234-446-797
Sequence 797, Application US/60234446
GENERAL INFORMATION:

APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
FILE REFERENCE: CLO00832
CURRENT APPLICATION NUMBER: US/60/234.446
CURRENT FILING DATE: 2000-09-18
NUMBER OF SEQ ID NOS: 1797
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 797
LENGTH: 71
TYPE: PRT
ORGANISM: HUMAN

Query Match 69.8%; Score 393.5; DB 15; Length 76;

US-09-110-716-40

US-09-110-716-40

US-09-110-716-40

US-60-234-446-797

Query Match 60.88; Score 343; DB 26; Length 71;
 Best Local Similarity 97.18; Pred. No. 1.3e-31;
 Matches 68; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 14 MKLLMVLLAALLHCTYADSGCKLLIEDVVEKTIINSISIPYKELLOEFIDSDAAAEAMG 73
 DB 1 MKLLMVLLAALLHCTYADSGCKLLIEDVVEKTIINSISIPYKELLOEFIDSDAAAEAMG 60

QY 74 KFKQCFILNQ 83
 DB 61 KFKQCFILNQ 70

Search completed: June 20, 2002, 11:05:26
 Job time: 222 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 20, 2002, 11:01:44 ; Search time 12.89 Seconds
(without alignments)
204,652 Million cell updates/sec

Title: US-09-673-395A-238

Perfect score: 564
Sequence: 1 CHARLNTDSSRLAKMLNL.....NFGIMHTVDSIMCNKSN 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496	87.9	95	3	US-08-821-451A-6
2	496	87.9	95	4	US-09-263-810-6
3	496	87.9	95	4	US-09-583-169-6
4	281	49.8	93	1	US-08-455-896-2
5	281	49.8	93	2	US-08-933-149-2
6	281	49.8	93	2	US-08-253-2
7	281	49.8	93	3	US-09-082-253-2
8	281	49.8	93	5	PCT-US96-08235-2
9	196.5	34.8	95	3	US-08-821-451A-27
10	196.5	34.8	95	4	US-09-263-810-27
11	196.5	34.8	95	4	US-09-583-169-27
12	195.5	34.7	95	1	US-08-455-896-7
13	195.5	34.7	95	2	US-08-933-149-7
14	195.5	34.7	95	2	US-09-082-253-7
15	195.5	34.7	95	2	US-09-082-253-7
16	195.5	34.7	95	5	PCT-US96-08235-7
17	65	11.5	1255	3	US-08-947-823-3
18	63.5	11.3	582	1	US-08-431-080-16
19	63.5	11.3	582	2	US-08-938-534-16
20	61.5	10.9	1604	2	US-09-004-838-95
21	61	10.8	445	2	US-08-900-148-2
22	61	10.8	446	2	US-08-922-171-3
23	61	10.8	472	2	US-08-922-171-2
24	61	10.8	1257	3	US-08-947-823-5
25	60.5	10.7	971	3	US-09-112-450-2
26	60.5	10.7	2471	3	US-09-112-450-4
27	59.5	10.5	114	1	US-08-031-399-3

28	59.5	10.5	114	1	US-08-031-399-6	Sequence 6, Appl
29	59.5	10.5	114	1	US-08-031-399-12	Sequence 12, Appl
30	59.5	10.5	114	1	US-08-393-305-3	Sequence 3, Appl
31	59.5	10.5	114	1	US-08-393-305-6	Sequence 6, Appl
32	59.5	10.5	114	1	US-08-726-817-3	Sequence 3, Appl
33	59.5	10.5	114	1	US-08-726-817-6	Sequence 6, Appl
34	59.5	10.5	114	1	US-08-504-042-3	Sequence 3, Appl
35	59.5	10.5	114	1	US-08-504-042-6	Sequence 6, Appl
36	59.5	10.5	114	1	US-08-504-042-12	Sequence 12, Appl
37	59.5	10.5	114	2	US-08-725-969-3	Sequence 3, Appl
38	59.5	10.5	114	2	US-08-725-969-6	Sequence 6, Appl
39	59.5	10.5	114	2	US-08-794-524-3	Sequence 3, Appl
40	59.5	10.5	114	2	US-08-794-524-6	Sequence 6, Appl
41	59.5	10.5	114	4	US-09-189-193-3	Sequence 3, Appl
42	59.5	10.5	114	4	US-09-189-193-6	Sequence 6, Appl
43	59.5	10.5	114	5	PCT-US94-03793-3	Sequence 3, Appl
44	59.5	10.5	114	5	PCT-US94-03793-6	Sequence 6, Appl
45	59.5	10.5	114	5	PCT-US94-03793-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-821-451A-6
; Sequence 6, Application US/08821451A
; Patent No. 6066724
; GENERAL INFORMATION:
; APPLICANT: Jian NL, Guo-Liang Yu and Reiner Gentz
; TITLE OF INVENTION: Human Endometrial Specific Steroid-
; TITLE OF INVENTION: Binding Factor I, II and III
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CECCHI, STEFANE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,451A
; FILING DATE: March 21, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,724
; FILING DATE: March 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-521 (PR257)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SFO ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-821-451A-6

Query Match 87.9%; Score 496; DB 3; Length 95;
Best Local Similarity 100.0%; Pred. No. 8,8e-54;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,896
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-455-896-2

Query Match 49.8%; Score 281; DB 1; Length 93;
Best Local Similarity 58.7%; Pred. No. 2,4e-27;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Caps 0;

QY 14 MKLWVLMALAILHGYAAGSCCKLEDEWEKTIINSDISIPYKELQRFIDSDAAAPAMG 73
DB 1 MKLWVLMALAILHGYAAGSCCKLEDEWEKTIINSDISIPYKELQRFIDSDAAAPAMG 60
QY 74 KFKQCFINOSHRTIKNFGIMHMTYVDSIWCNM 105
DB 61 ELKECFINQDTFTLSNVEVFQMLIYDSSLCDL 92

RESULT 5
US-08-933-149-2
Sequence 2, Application US/08933149
Patent No. 5922836
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: MAMMAGLOBIN, A SECRETED
TITLE OF INVENTION: MAMMARY SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,149
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: HENDERSON, MELODIE W.
REGISTRATION NUMBER: 37,848
REFERENCE/DOCKET NUMBER: 6029-6040
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-933-149-2

Query Match 49.8%; Score 281; DB 2; Length 93;
Best Local Similarity 58.7%; Pred. No. 2,4e-27;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Caps 0;

QY 14 MKLWVLMALAILHGYAAGSCCKLEDEWEKTIINSDISIPYKELQRFIDSDAAAPAMG 73
DB 1 MKLWVLMALAILHGYAAGSCCKLEDEWEKTIINSDISIPYKELQRFIDSDAAAPAMG 60
QY 74 KFKQCFINOSHRTIKNFGIMHMTYVDSIWCNM 105
DB 61 ELKECFINQDTFTLSNVEVFQMLIYDSSLCDL 92

RESULT 6
US-09-082-343-2
Sequence 2, Application US/09082343
Patent No. 5968754
GENERAL INFORMATION:
APPLICANT: WATSON, MARK A.
APPLICANT: FLEMING, TIMOTHY P.
TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105-1817
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,343
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/455,896
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 952726
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-082-343-2

Query Match 49.8%; Score 281; DB 2; Length 93;
Best Local Similarity 58.7%; Pred. No. 2,4e-27;
Matches 54; Conservative 13; Mismatches 25; Indels 0; Caps 0;

```

OY      14  MKLIMVMTLALLHCYAAAGCKLLEDVETKINSIDSIPEYKELLOEFDPSAAAEAG  73
OY      14  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1  MKLIMVMTLALLHCYAAAGCKPLENTISKTIINQYASKTEYKELLOEFDIDNATNAID  60
OY      74  KFKOCFLNOSHRLTKNFGMLMTIYVDSIMCN  105
OY      61  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61  ELKKEFLNODTELISNVEFMOLITDSSICDL  92

```

```

1      RESULT 7
2      US-09-082-253-2
3      ; Sequence 2, Application US/09082253
4      ; Patent No. 6004756
5      GENERAL INFORMATION:
6      APPLICANT: WATSON, MARK A.
7      APPLICANT: FLEMING, TIMOTHY P.
8      TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
9      TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
10     NUMBER OF SEQUENCES: 13
11     CORRESPONDENCE ADDRESS:
12     ADDRESSEE: ROGERS, HOWELL & HAFERKAMP
13     STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
14     CITY: ST. LOUIS
15     STATE: MISSOURI
16     COUNTRY: USA
17     ZIP: 63105-1817
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: Patent In Release #1.0, Version #1.25
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/09/082,253
25     FILING DATE:
26     CLASSIFICATION:
27     PRIOR APPLICATION DATA:
28     APPLICATION NUMBER: 08/455,896
29     FILING DATE: 05/31/1995
30     ATTORNEY/AGENT INFORMATION:
31     NAME: HOLLAND, DONALD R.
32     REGISTRATION NUMBER: 35,197
33     REFERENCE/DOCKET NUMBER: 953726
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE: (314) 727-5188
36     TELEFAX: (314) 727-6092
37     INFORMATION FOR SEQ ID NO: 2:
38     SEQUENCE CHARACTERISTICS:
39     LENGTH: 93 amino acids
40     TYPE: amino acid
41     STRANDEDNESS: single
42     TOPOLOGY: linear
43     MOLECULE TYPE: protein
44     HYPOTHETICAL: NO
45     ;
46     ;
47     ;
48     ;
49     ;
50     ;
51     ;
52     ;
53     ;
54     ;
55     ;
56     ;
57     ;
58     ;
59     ;
60     ;
61     ;
62     ;
63     ;
64     ;
65     ;
66     ;
67     ;
68     ;
69     ;
70     ;
71     ;
72     ;
73     ;
74     ;
75     ;
76     ;
77     ;
78     ;
79     ;
80     ;
81     ;
82     ;
83     ;
84     ;
85     ;
86     ;
87     ;
88     ;
89     ;
90     ;
91     ;
92     ;
93     ;
94     ;
95     ;
96     ;
97     ;
98     ;
99     ;
100    ;
101    ;
102    ;
103    ;
104    ;
105    ;
106    ;
107    ;
108    ;
109    ;
110    ;
111    ;
112    ;
113    ;
114    ;
115    ;
116    ;
117    ;
118    ;
119    ;
120    ;
121    ;
122    ;
123    ;
124    ;
125    ;
126    ;
127    ;
128    ;
129    ;
130    ;
131    ;
132    ;
133    ;
134    ;
135    ;
136    ;
137    ;
138    ;
139    ;
140    ;
141    ;
142    ;
143    ;
144    ;
145    ;
146    ;
147    ;
148    ;
149    ;
150    ;
151    ;
152    ;
153    ;
154    ;
155    ;
156    ;
157    ;
158    ;
159    ;
160    ;
161    ;
162    ;
163    ;
164    ;
165    ;
166    ;
167    ;
168    ;
169    ;
170    ;
171    ;
172    ;
173    ;
174    ;
175    ;
176    ;
177    ;
178    ;
179    ;
180    ;
181    ;
182    ;
183    ;
184    ;
185    ;
186    ;
187    ;
188    ;
189    ;
190    ;
191    ;
192    ;
193    ;
194    ;
195    ;
196    ;
197    ;
198    ;
199    ;
200    ;
201    ;
202    ;
203    ;
204    ;
205    ;
206    ;
207    ;
208    ;
209    ;
210    ;
211    ;
212    ;
213    ;
214    ;
215    ;
216    ;
217    ;
218    ;
219    ;
220    ;
221    ;
222    ;
223    ;
224    ;
225    ;
226    ;
227    ;
228    ;
229    ;
230    ;
231    ;
232    ;
233    ;
234    ;
235    ;
236    ;
237    ;
238    ;
239    ;
240    ;
241    ;
242    ;
243    ;
244    ;
245    ;
246    ;
247    ;
248    ;
249    ;
250    ;
251    ;
252    ;
253    ;
254    ;
255    ;
256    ;
257    ;
258    ;
259    ;
260    ;
261    ;
262    ;
263    ;
264    ;
265    ;
266    ;
267    ;
268    ;
269    ;
270    ;
271    ;
272    ;
273    ;
274    ;
275    ;
276    ;
277    ;
278    ;
279    ;
280    ;
281    ;
282    ;
283    ;
284    ;
285    ;
286    ;
287    ;
288    ;
289    ;
290    ;
291    ;
292    ;
293    ;
294    ;
295    ;
296    ;
297    ;
298    ;
299    ;
300    ;
301    ;
302    ;
303    ;
304    ;
305    ;
306    ;
307    ;
308    ;
309    ;
310    ;
311    ;
312    ;
313    ;
314    ;
315    ;
316    ;
317    ;
318    ;
319    ;
320    ;
321    ;
322    ;
323    ;
324    ;
325    ;
326    ;
327    ;
328    ;
329    ;
330    ;
331    ;
332    ;
333    ;
334    ;
335    ;
336    ;
337    ;
338    ;
339    ;
340    ;
341    ;
342    ;
343    ;
344    ;
345    ;
346    ;
347    ;
348    ;
349    ;
350    ;
351    ;
352    ;
353    ;
354    ;
355    ;
356    ;
357    ;
358    ;
359    ;
360    ;
361    ;
362    ;
363    ;
364    ;
365    ;
366    ;
367    ;
368    ;
369    ;
370    ;
371    ;
372    ;
373    ;
374    ;
375    ;
376    ;
377    ;
378    ;
379    ;
380    ;
381    ;
382    ;
383    ;
384    ;
385    ;
386    ;
387    ;
388    ;
389    ;
390    ;
391    ;
392    ;
393    ;
394    ;
395    ;
396    ;
397    ;
398    ;
399    ;
400    ;
401    ;
402    ;
403    ;
404    ;
405    ;
406    ;
407    ;
408    ;
409    ;
410    ;
411    ;
412    ;
413    ;
414    ;
415    ;
416    ;
417    ;
418    ;
419    ;
420    ;
421    ;
422    ;
423    ;
424    ;
425    ;
426    ;
427    ;
428    ;
429    ;
430    ;
431    ;
432    ;
433    ;
434    ;
435    ;
436    ;
437    ;
438    ;
439    ;
440    ;
441    ;
442    ;
443    ;
444    ;
445    ;
446    ;
447    ;
448    ;
449    ;
450    ;
451    ;
452    ;
453    ;
454    ;
455    ;
456    ;
457    ;
458    ;
459    ;
460    ;
461    ;
462    ;
463    ;
464    ;
465    ;
466    ;
467    ;
468    ;
469    ;
470    ;
471    ;
472    ;
473    ;
474    ;
475    ;
476    ;
477    ;
478    ;
479    ;
480    ;
481    ;
482    ;
483    ;
484    ;
485    ;
486    ;
487    ;
488    ;
489    ;
490    ;
491    ;
492    ;
493    ;
494    ;
495    ;
496    ;
497    ;
498    ;
499    ;
500    ;
501    ;
502    ;
503    ;
504    ;
505    ;
506    ;
507    ;
508    ;
509    ;
510    ;
511    ;
512    ;
513    ;
514    ;
515    ;
516    ;
517    ;
518    ;
519    ;
520    ;
521    ;
522    ;
523    ;
524    ;
525    ;
526    ;
527    ;
528    ;
529    ;
530    ;
531    ;
532    ;
533    ;
534    ;
535    ;
536    ;
537    ;
538    ;
539    ;
540    ;
541    ;
542    ;
543    ;
544    ;
545    ;
546    ;
547    ;
548    ;
549    ;
550    ;
551    ;
552    ;
553    ;
554    ;
555    ;
556    ;
557    ;
558    ;
559    ;
560    ;
561    ;
562    ;
563    ;
564    ;
565    ;
566    ;
567    ;
568    ;
569    ;
570    ;
571    ;
572    ;
573    ;
574    ;
575    ;
576    ;
577    ;
578    ;
579    ;
580    ;
581    ;
582    ;
583    ;
584    ;
585    ;
586    ;
587    ;
588    ;
589    ;
590    ;
591    ;
592    ;
593    ;
594    ;
595    ;
596    ;
597    ;
598    ;
599    ;
600    ;
601    ;
602    ;
603    ;
604    ;
605    ;
606    ;
607    ;
608    ;
609    ;
610    ;
611    ;
612    ;
613    ;
614    ;
615    ;
616    ;
617    ;
618    ;
619    ;
620    ;
621    ;
622    ;
623    ;
624    ;
625    ;
626    ;
627    ;
628    ;
629    ;
630    ;
631    ;
632    ;
633    ;
634    ;
635    ;
636    ;
637    ;

```

```

Query Match Similarity      49.8% Score 281; DB 3; Length 93;
Best Local Similarity      58.7%; Pred. No. 2,4e+27;
Matches 54; Conservative   13; Mismatches 25; Indels    0; Gaps    0;

OY      14 MKLIMVLMIALLLHCYADSGCKLEDMEYEKTI NSDI SIPEYKEKLQEFITDSDAALAEAMC 73
          | | | | | | | | | | | | | | : | | | | | | | | : |
Db       1 MKLIMVLMIALLSQHCYAGS GCPLEENYSIKTINPVSSTKEYEKLQEFIDNATTAID 60
          | | | | | | | | | | | | | | : | | | | | | | | : |

OY      74 KFKQCFLNOSHRTLKRFGLMHTIYVDISNCNM 105
          : | | | | | | | | : | : | | | | | : :
Db       61 ELKECFLNQTDFTLSNVFYMOLLIYDSSICDL 92

RESULT      8
PCT-US96--08235-2
; Sequence 2, Application PC/TUS9608235
; GENERAL INFORMATION:
```

APPLICANT: MATSON, MARK A.
 APPLICANT: FLEMING, TIMOTHY P.
 TITLE OF INVENTION: DNA SEQUENCE AND ENCODED
 TITLE OF INVENTION: MAMMARY-SPECIFIC BREAST CANCER PROTEIN
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ROGERS, HOWELL, & HAFERKAMP
 STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
 CITY: ST. LOUIS
 STATE: MISSOURI
 COUNTRY: USA
 ZIP: 63105-1817
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/08235
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 964796
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (314) 727-5188
 TELEFAX: (314) 727-6092
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 93 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO

Query Match	49.8%;	Score 281;	DB 5;	Length 93;
Best Local Similarity	58.7%;	Pred. 204-27;		
Matches	54;	Conservative 13;	Mismatches 25;	Indels 0;
				Gaps 0;
QY	14	MLLMLMLMALALHCAVDSGCKLLIDYNEKTIINSISIPYKKELLOFIDSDAARFAMC	73	
DB	1	MLLMLMLMALALSHCAAGSGCPLELVNISKRTINPQVSKTEKKELLOFIDDNATTAID	60	
QY	74	KFKQCFHINQSHRTILKNFGLMMHTVYYSIMCMM	105	
DB	61	ELKKELFLNQTDETLISNVKFFPOLITVDSGLD	92	

RESULT 9
 US-08-821-451A-27
 ; Sequence 27, Application US/08821451A
 ; Patent No. 6066724
 ; GENERAL INFORMATION:
 ; APPLICANT: Jian NI, Guo-Liang Yu and Reiner Gentz
 ; TITLE OF INVENTION: Human Endometrial Specific Steroid
 ; TITLE OF INVENTION: Binding Factor 1, 11 and 111
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
 ; STREET: 6 BECKER FARM ROAD
 ; CITY: ROSELAND
 ; STATE: NEW JERSEY
 ; COUNTRY: USA
 ; ZIP: 07068
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS

```

SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,451A
FILING DATE: March 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,724
FILING DATE: March 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (P257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-821-451A-27

```

```

Query Match Similarity      34.8%; Score 196.5; DB 3; Length 95;
Best Local Similarity      35.8%; Pred. No. Ge-17; Mismatches 29; Indels 1; Gaps 1
Matches      34; Conservative      31;

QY      14 MKLWVIAIALLLCYAD-SGCKLELDWEKTVNSDISIPXEKELLOPFIIDSDAAAEAM 72
       111::: :|:: ||| ||| :::: |||| :::: |:: :::: :|:
DB      1 MKLVFLLVTTPICCYAGSCSCLDEVIKGTINSTVTLHDYMKLVKPYOAHFTKEAV 60

QY      73 GKFKCFLNQSHRTLKNGLMHTYYDSTWCNKS 107
       :||||::: :||:|:|| :::-|-|
DB      61 KQKCCFLDQTDLTKLENVGMMALFNSESCOOPS 95

RESULT 10
US-09-263-810-27
; Sequence 27, Application US/09263810
; Patent No. 6174992
GENERAL INFORMATION:
APPLICANT: Jian Ni, Guo-liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
TITLE OF INVENTION: Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSPE: BECCI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,810
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821,451
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 32500-521 (PF257)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700

```

```

: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 95 AMINO ACIDS
:   TYPE: AMINO ACID
:   STRANDEDNESS: SINGLE
:   TOPOLOGY: LINEAR
:   MOLECULE TYPE: PROTEIN
US-09-263-810-27

```

	Best Match	34.8%	Score 196.5;	DB 4;	Length 95;
	Query Local Similarity	35.8%	Pred. No. 6e-17;		
	Matches	34;	Conservative	31;	Mismatches 29; Indels 1; Gaps 1;
Oy	14	MLKLIVLMALLHICVAD-SGCKLEIDEMVEKTIISDSIFSEYKELOEPIIDSDAAAEAM	72		
		::: ::: ::: ::: ::: ::: :::			
Dd	1	MKLVFELVLVTPTICTAASGSGSILDEXIKRTINSTYTLLMDYAKI.VKPVQAHLTKENAV	60		
Oy	73	GKFKOCFLNOSHRIPLKNFGIMHTTVTVDSTWCMNKS	107		
		::: ::: ::: ::: ::: ::: ::: :::			
Dd	61	KQFKQCFLDITDKTLLENVGVMALINSSCOOPS	95		

```

11 RESULT
12 US-09-583-169-27
13 ; Sequence 27, Application US/09583169
14 ; Patent No. 6338948
15 ; GENERAL INFORMATION:
16 ; APPLICANT: Jian Ni, Guo-liang Yu and Reiner Gentsch
17 ; TITLE OF INVENTION: Human Endometrial Specific Steroid-
18 ; TITLE OF INVENTION: Binding Factor I, II and III
19 ; NUMBER OF SEQUENCES: 27
20 ; CORRESPONDENCE ADDRESS:
21 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GIUFFILAN,
22 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
23 ; STREET: 6 BECKER FARM ROAD
24 ; CITY: ROSELAND
25 ; STATE: NEW JERSEY
26 ; COUNTRY: USA
27 ; ZIP: 07068
28 ; COMPUTER READABLE FORM:
29 ; MEDIUM TYPE: 3.5 INCH DISKETTE
30 ; COMPUTER: IBM PS/2
31 ; OPERATING SYSTEM: MS-DOS
32 ; SOFTWARE: WORD PERFECT 5.1
33 ; CURRENT APPLICATION DATA:
34 ; APPLICATION NUMBER: US/09/583,169
35 ; FILING DATE:
36 ; CLASSIFICATION:
37 ; PRIOR APPLICATION DATA:
38 ; APPLICATION NUMBER: 08/821,451
39 ; FILING DATE:
40 ; ATTORNEY/AGENT INFORMATION:
41 ; NAME: MULLINS, J.G.
42 ; REGISTRATION NUMBER: 33,073
43 ; REFERENCE/DOCKET NUMBER: 325800-521 (PF257)
44 ; TELECOMMUNICATION INFORMATION:
45 ; TELEPHONE: 201-994-1700
46 ; TELEFAX: 201-994-1744
47 ; INFORMATION FOR SEQ ID NO: 27:
48 ; SEQUENCE CHARACTERISTICS:
49 ; LENGTH: 95 AMINO ACIDS
50 ; TYPE: AMINO ACID
51 ; STRANDEDNESS: SINGLE
52 ; TOPOLOGY: LINEAR
53 ; MOLECULE TYPE: PROTEIN
54 ;
55 US-09-583-169-27
56
57 Query Match 34.8%; Score 196.5; DB 4; Length 95;
58 Best Local Similarity 35.8%; Pred. No. 6e-17; 29; Indels 1; Gaps 1;
59 Matches 34; Conservative 31; Mismatches 31;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2002, 11:01:44 ; Search time 30.14 Seconds
(without alignments)
398.008 Million cell updates/sec

Title: US-09-673-395a-238
Perfect score: 564
Sequence: 1 CHARLNTDSSRLANKLMLV.....NFGIMHTVYDSICWKNKS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_032802:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	564	100.0	108	20	AAV60038 Human endometrium
2	496	87.9	95	18	AAV35804 Human endometrial
3	496	87.9	95	20	AAV02590 Human mamaglobin
4	496	87.9	95	21	AAV03769 Human endometrial
5	496	87.9	95	21	AAV92226 Human endometrial
6	496	87.9	95	21	AAV92237 Mamaglobin homolo
7	496	87.9	95	21	AAV65394 Human 5' EST relat
8	496	87.9	95	22	AAV31682 An endometrial spe
9	390	69.1	74	19	AAV61649 Non-ocular disease
10	281	49.8	93	18	AAV10179 Mammary-specific s
11	281	49.8	93	19	AAV59777 Amino acid sequenc

12	281	49.8	93	19	AAV48432 Mamaglobin protei
13	281	49.8	93	20	AAV01718 Mamaglobin, a mam
14	281	49.8	93	21	AAV13786 Human mamaglobin
15	281	49.8	93	21	AAV84622 Amino acid sequenc
16	281	49.8	93	22	AAV07517 Human mamaglobin
17	281	49.8	93	22	AAV51127 Human mamaglobin
18	281	49.8	410	22	AAV33359 Human breast cance
19	281	49.8	743	22	AAV33358 Human breast cance
20	281	49.8	1095	22	AAV33357 Human breast cance
21	280	49.6	93	22	AAV07531 Human mamaglobin
22	279	49.5	93	22	AAV07529 Human mamaglobin
23	277	49.1	93	22	AAV07534 Human mamaglobin
24	277	49.1	93	22	AAV07535 Human mamaglobin
25	277	49.1	93	22	AAV07536 Human mamaglobin
26	273	48.4	93	22	AAV07530 Human mamaglobin
27	272	48.2	93	22	AAV07532 Human mamaglobin
28	269	47.7	93	22	AAV07533 Human mamaglobin
29	265.5	47.1	90	22	AAV07528 Human mamaglobin
30	265.5	47.1	90	22	AAV07537 Human mamaglobin
31	206	36.5	74	21	AAV44624 Amino acid sequenc
32	95	16.8	33	19	AAV48433 Mamaglobin synthe
33	77	13.7	21	22	AAV51114 Human mamaglobin
34	77	13.7	21	22	AAV51124 Mamaglobin monocl
35	77	13.7	21	22	AAV51126 Mamaglobin synthe
36	77	13.7	30	19	AAV48435 Drosophila melano
37	73	12.9	702	22	ABV63821 Human mamaglobin
38	72	12.8	20	22	AAV51132 Drosophila melano
39	71	12.6	828	22	ABV62649 Drosophila melano
40	71	12.6	1187	22	ABV63669 Drosophila melano
41	70.5	12.5	343	22	ABV64633 G protein-coupled
42	70	12.4	457	22	AAV03813 Drosophila melano
43	68.5	12.1	384	22	ABV69698 Human colon carcin
44	68.5	12.1	619	22	AAV97025 Human mamaglobin
45	68	12.1	55	22	AAV51128 Mamaglobin

ALIGNMENTS

RESULT 1
ID AAV60038 standard: Protein; 108 AA.
AAV60038;
AC AAV60038;
XX 31-JAN-2000 (first entry)
DT Human endometrium tumour EST encoded protein 98.
DE Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
KW treatment; uterine; gene therapy; expressed sequence tag.
XX Homo sapiens.
OS DE19817948-A1.
PN 21-OCT-1999.
PD 17-APR-1998; 98DE-1017948.
PF 17-APR-1998; 98DE-1017948.
PR 17-APR-1998; 98DE-1017948.
XX (META-) METAGEN GDS GENOMFORSCHUNG MBH.
PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
PI WPI: 1999-591957/51.
DR N-PSDB; AA42013.
XX New nucleic acid sequences expressed in uterine cancer tissues, and
PT derived polypeptides, for treatment of uterine and endometrial cancer
PT and identification of therapeutic agents -

PS Claim 23: Page 314: 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for
 CC generation of specific antibodies. (A) are identified by assembling ESTs
 CC (expressed sequence tags) from a particular tissue type before comparison
 CC of expression patterns. This allows a significantly longer fragment of
 CC the gene to be revealed, so should reduce the number of failures may represent
 CC associated with the fact that ESTs from different libraries may represent
 CC different parts of the same unknown gene, distorting the estimated
 CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent
 CC protein fragments encoded by the human endometrium tumour cDNA library
 CC derived EST fragments represented in AA41981-442121.

SO Sequence 108 AA:

Query Match 100.0%; Score 564; DB 20; Length 108;
 Best Local Similarity 100.0%; Pred. No. 1.5e-53; Indels 0; Gaps 0;
 Matches 108; Conservative 0; Mismatches 0;

OY 1 CHARLNTDSRLAKMLVLMALALHCVADSGCKLEDMVEKINSISIPYKELQ 60
 Db 1 charlntdsrlamkllmvlmlaalhhcyadsgcklledmvektinsdisipykellq 60
 OY 61 EFIDSDAAAEAMGKFKOCFLNOSHRTLNFGIMHTVYDSIWCNMKSN 108
 Db 61 efidsdaaeamgkfkqclnshrtlnfgimhtvydsiwcnmksn 108

RESULT 2

AAW35804 standard; Protein: 95 AA.

AAW35804;

27-MAR-1998 (first entry)

Human endometrial specific steroid-binding factor III.

Endometrial specific steroid-binding factor III; ESR III; human;

Clara cell secretory protein; endometrium;

phospholipase A2 inhibitor; polychlorinated biphenyl; antiagregant;

Inflammation; asthma; rhinitis; cystic fibrosis; airway disease;

neoplasia; atopy; therapy; diagnosis.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..21 /label= Sig_peptide

Protein 22..95 /label= Mat_protein

WO9734997-A1.

25-SEP-1997.

21-MAR-1996; 96WO-US03857.

21-MAR-1996; 96WO-US03857.

(HUMA-) HUMAN GENOME SCI INC.

Gentz RL, Ni J, Yu G;

WPI: 1997-480206/44.

N-PSDB: AAT94832.

DR

XX Human endometrial specific steroid-binding factor I, II and III -
 PT used to treat inflammation, asthma, rhinitis, cystic fibrosis,
 PT airway disease, neoplasia, atopy etc.

PS Claim 19: Page 65; 92pp; English.

XX This sequence comprises human endometrial specific steroid binding
 CC factor III (ESF III), a protein that inhibits phospholipase A2
 CC activity, binds to polychlorinated biphenyl compounds, reduces
 CC foreign protein antigenicity, inhibits monocyte and neutrophil
 CC chemotaxis and phagocytosis, inhibits platelet aggregation,
 CC regulates eicosanoid levels in the human uterus and controls the
 CC growth of endometrial cells. The amino acid sequence was deduced
 CC from a cDNA clone (see AAT94832) and ESF II (see AAW35803) are also
 CC claimed. Human ESF III has about 36% identity with rat prostatic
 CC steroid-binding protein. Recombinant ESF I, II and III can be
 CC expressed in host cells for use in claimed methods (a) for treating
 CC a patient in need of ESF I, II or III (including compounds which bind
 CC to and inhibit activation of the ESF polypeptide, hESF I, II and
 CC III may be used to treat inflammation, asthma, rhinitis, cystic
 CC fibrosis, airway disease, neoplasia and atopy.

SO Sequence 95 AA:

Query Match 87.9%; Score 496; DB 18; Length 95;
 Best Local Similarity 100.0%; Pred. No. 2.9e-46; Indels 0; Gaps 0;
 Matches 95; Conservative 0; Mismatches 0;

OY 14 MKLIMVLMALALHCVADSGCKLEDMVEKINSISIPYKELQEFIDSDAAEAMG 73
 Db 1 mklmvlmlaalhhcyadsgcklledmvektinsdisipykellqefidsdaaeamg 73

OY 74 KFKOCFLNOSHRTLNFGIMHTVYDSIWCNMKSN 108
 Db 61 kfkqclnshrtlnfgimhtvydsiwcnmksn 95

RESULT 3

AAW02590 standard; Protein: 95 AA.

AAW02590;

26-JUL-1999 (first entry)

A human mamaglobin homologue (HMH).

Human mamaglobin homologue; HMH; antagonist; neoplastic disorder;

adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;

teratocarcinoma; endometriosis.

Homo sapiens.

WO9919487-A1.

22-APR-1999.

14-OCT-1998; 98WO-US21729.

16-OCT-1997; 97US-0951750.

(INCY-) INCYTE PHARM INC.

Hillman JL, Murry LE, Shah P;

WPI: 1999-302531/25.

N-PSDB: AAX36138.

DR

New human mamaglobin homologue (HMH), useful for diagnosing, treating

PT

PT Diagnosing, staging and monitoring gynecological cancer comprising
 PT using an elevated level of ESBP111 in a patient as an indicator of
 PT cancer

PS Claim 6; Page 28-29; 32pp; English.

CC The levels of human endometrial specific steroid binding factor
 CC (ESBP111) can be measured and compared to control levels and used to
 CC diagnose the presence of a gynaecological (uterine, breast, endometrial,
 CC or ovarian) cancer in a patient. ESBP111 levels can also be used to
 CC diagnose metastasis, to stage or monitor gynaecological cancer.
 CC Antibodies specific for ESBP111 can be used to treat gynecological
 CC cancers.

XX Sequence 95 AA;

Query Match 87.9%; Score 496; DB 21; Length 95;

Best Local Similarity 100.0%; Pred. No. 2.9e-46; Mismatches 0; Indels 0; Gaps 0;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

14 MKLWMLAALLHCYADSGCKLEDMVEKTI NSDISIPEYKELLQEFIDSDAAEAMG 73

1 mklmvlmaalllhcycadsgcklleedmvektinsdisipeykellqefidsdaaamg 60

74 KFKQCFINQSHRTLNKFGIMHFTVYDSIMCNMKN 108

61 kfkqcfinsghrtlnknfglmhftvdsimcnmkn 95

RESULT 6

AAV92237 standard; Protein; 95 AA.

AAV92237;

10-AUG-2000 (first entry)

Mammoglobin homologue from clone Mamm-X.

XX Clone Mamm-X; mammoglobin; breast cancer; cytostatic; anti-HIV;
 XX immunosuppressive; anti-allergic; anti-infective; anti-inflammatory;
 XX anti-arteritic; anti-arteriosclerotic; vasotrophic; neuroprotective;
 XX neurotrophic; dermatological; tranquilizer; vulnery.

OS Homo sapiens.

PN WO200020447-A2.

PD 13-APR-2000.

PF 06-OCT-1999; 99WO-US22294.

PR 06-OCT-1998; 98US-0103195.

PR 05-OCT-1999; 99US-0103195.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA;

DR WPI; 2000-303741/26.

DR N-PSDB; AAA09118.

PT Nucleic acids encoding polypeptides with syncollin-like, claudin-like or
 PT cytochrome-like activity, useful for treating diseases including cancer,
 PT Alzheimer's and atherosclerosis

PS Claim 23; Fig 6; 118pp; English.

CC Clone Mamm-X encodes a polypeptide that is 100 percent identical to human
 CC Mammoglobin B precursor, a potential marker of breast cancer nodal
 CC metastasis. The sequences are useful for treatment of diseases such as
 CC cancer, immune disorders, autoimmune disease, acquired immune deficiency

CC syndrome (AIDS), transplant rejection, allergy, infection by a
 CC pathological agent or organism, inflammatory disorders, arthritis, a
 CC haematopoietic disorder, a skin disorder, atherosclerosis, restenosis, a
 CC neurological disease, Alzheimer's disease, trauma, spinal cord injury
 CC and skeletal disorders.

XX Sequence 95 AA;

Query Match 87.9%; Score 496; DB 21; Length 95;

Best Local Similarity 100.0%; Pred. No. 2.9e-46; Mismatches 0; Indels 0; Gaps 0;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

14 MKLWMLAALLHCYADSGCKLEDMVEKTI NSDISIPEYKELLQEFIDSDAAEAMG 73

1 mklmvlmaalllhcycadsgcklleedmvektinsdisipeykellqefidsdaaamg 60

74 KFKQCFINQSHRTLNKFGIMHFTVYDSIMCNMKN 108

61 kfkqcfinsghrtlnknfglmhftvdsimcnmkn 95

RESULT 7

AAV65394 standard; Protein; 95 AA.

AAV65394;

01-FEB-2000 (first entry)

XX Human 5' EST related polypeptide SEQ ID NO:1555.
 XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 XX gene therapy; chromosome mapping; upstream regulatory sequence;
 XX forensic; location; development; protein synthesis; stability;
 XX regulation; identification.

OS Homo sapiens.

PN WO9953051-A2.

PD 21-OCT-1999.

PF 09-APR-1999; 99WO-1H00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

PA (GEST) GENSET.

PI Dumas Mline Edwards J, Duclert A, Giordano J;

DR WPI; 2000-038446/03.

DR N-PSDB; AA243008.

PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

PS Claim 3; Page 818; 837pp; English.

CC AA242265 to AA243075 represent novel 5' expressed sequence tag (5ST)
 CC sequences, corresponding to human secreted proteins. AA164651 to
 CC AA165438 represent the EST-related proteins corresponding to AA242265 to
 CC AA243072. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be

Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 32 DSGCKLEDMVEKTSIDSIPEYKELDFIDSDAAAMKFKOCFLNOSHRTLKNG 91
 DB 1 dsgcklledmvektlndisidipeykellqefidsdaaamgkfkqclngshrtlkng 60

OY 92 LMMHTVYDSICNM 105
 DB 61 lmmhtvydsicnm 74

RESULT 10

AAW10179
 ID AAW10179 standard; Protein; 93 AA.

AC AAW10179;

DT 12-AUG-1997 (first entry)

DE Mammary-specific secretory protein, mammaglobin.

KW mammaglobin; mammary-secretory protein; breast cancer; detection;
 neoplastic disease; diagnosis.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..19 /label= signal_peptide

FT Protein 20..93 /label= mature_protein

PN WO9638463-A1.

PD 05-DEC-1996.

PE 31-MAY-1996; 96MO-US08235.

PR 31-MAY-1995; 95US-0455896.

PA (UNIV) UNIV WASHINGTON.

PI Fleming TP, Watson MA;

DR WPI; 1997-034299/03.

DR N-PSDB; AAT50925.

XX Nucleic acid encoding mammary-specific secretory protein,
 PT mammaglobin - used to develop prods. for the early diagnosis and
 PT treatment of breast cancer neoplastic disease

PS Claim 3; Fig 2; 54pp; English.

XX The present sequence is that of a mammary-specific secretory protein
 CC designated mammaglobin, which is overexpressed in 27% of stage I primary
 CC breast cancer tumours. The anonymous sequence tag previously designated
 CC DSTR002 was used to demonstrate that mammaglobin is abundant in the
 CC breast cancer tumour cell line MDA-MB-415. To isolate the full-length
 CC mammaglobin cDNA (AAT50925), the mRNA was reverse transcribed from
 CC this cell line and cloned using the RACE PCR technique. The nucleic acid
 CC and protein can be used to develop prods. e.g. antibodies or probes, for
 CC the detection and treatment of breast neoplastic disease.

XX Sequence 93 AA;

Query Match 49.8%; Score 281; DB 18; Length 93;

Best Local Similarity 58.7%; Pred. No. 5.7e-23;

Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

14 MKLIMVIMLALLLHCTADSGCKLEDMVEKTSIDSIPEYKELDFIDSDAAAMK 73
 mklmvlmalaalshqcyagscpllenvisklnpyskylqellqefidnatnaid 60

OY 74 KFKOCFLNOSHRTLEKFGIAMHTVYDSICNM 105
 DB 61 kfkocflnqdetlshvevfmglydsslccl 92

RESULT 11

AAW59777
 ID AAW59777 standard; Protein; 93 AA.

AC AAW59777;

DT 12-OCT-1998 (first entry)

DE Amino acid sequence of the human steroid binding protein C2.

KW Human steroid-binding protein C2; hsbp2; hsbp1; breast cancer; probe;
 gene therapy vector; ribozyme; probe; hybridisation; amplification;
 antibody; immunoassay.

OS Homo sapiens.

PN WO9821331-A1.

PD 22-MAY-1998.

PE 07-NOV-1997; 97WO-US20674.

PR 12-NOV-1996; 96US-0747547.

PA (INCYT) INCYTE PHARM INC.

PI Akerblom IE, Goli SK, Hawkins PR, Hillman JL, Murry LE;

DR WPI; 1998-297935/26.

DR N-PSDB; AAV41580.

XX New human steroid binding proteins C1 and C2 - useful for, e.g.
 PT diagnosis, monitoring and treating breast cancer, and for drug
 PT screening

XX Claim 12; Fig 2; 70pp; English.

XX This is the amino acid sequence of the human steroid-binding protein
 CC C2 (hsbp2) used in the method of the invention for the diagnosis,
 CC monitoring and treatment of breast cancer. Hsbp1 and hsbp2 are useful
 CC as markers for breast cancer, i.e. measuring levels of hsbp1 and hsbp2
 CC used for diagnosis or monitoring the disease, to identify subjects
 CC at risk and to discriminate between different forms of cancer for
 CC selection of appropriate therapies. They may also be used for drug
 CC screening. Nucleic acids encoding hsbp1 and hsbp2 can be used in gene
 CC therapy vectors to over express the steroid-binding proteins, preventing
 CC binding of steroids, or antisense sequences, monitoring (by quantifying
 CC acids can also be used for the diagnosis and monitoring (by quantifying
 CC expression of hsbp), as source of probes for hybridisation and
 CC amplification of genomic or related sequences for studying regulation of
 CC gene function and for mapping the genomic sequence. Antibodies are used
 CC as diagnostic reagents in standard immunoassays for hsbp.

XX Sequence 93 AA;

Query Match 49.8%; Score 281; DB 19; Length 93;

Best Local Similarity 58.7%; Pred. No. 5.7e-23;

Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

OY 14 MKLIMVIMLALLLHCTADSGCKLEDMVEKTSIDSIPEYKELDFIDSDAAAMK 73
 DB 1 mklmvlmalaalshqcyagscpllenvisklnpyskylqellqefidnatnaid 60

OY 74 KFKOCFLNOSHRTLEKFGIAMHTVYDSICNM 105
 DB 61 kfkocflnqdetlshvevfmglydsslccl 92

RESULT 12

AAW48432

ID AAW48432 standard; Protein; 93 AA.

AC AAW48432;

DT 13-JUL-1998 (first entry)

XX Mammaglobin protein.

XX Mammaglobin; detection; diagnosis; breast cancer; tumour; antibody;

XX gene therapy; human.

XX Homo sapiens.

XX WO9807753-A1.

XX 26-FEB-1998.

XX 19-AUG-1997; 97WO-US14666.

XX 15-AUG-1997; 97US-0697106.

XX 19-AUG-1996; 96US-0697106.

XX (ABBO) ABBOTT LAB.

XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;

XX Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;

XX Roberts-Rapp L, Russell JC, Stroupe SD;

XX WPI; 1998-169096/15.

XX N-PSDB; AAV17905, AAV17906.

XX Antibodies to mammaglobin polypeptide(s) - used for detecting,

XX diagnosing, preventing or treating diseases or conditions of breast

XX such as breast cancer

XX Claim 8; Page 92; 105pp; English.

PS The present sequence represents mammaglobin which is used in an example
 XX of the present invention. The present invention describes an antibody
 CC (A) which specifically binds to at least 1 mammaglobin epitope (ME) which
 CC is derived from an amino acid sequence having at least 50% identity to
 CC an amino acid sequence (see AAW48432) and fragments. Also described are:
 CC (1) an assay kit for determining the presence of mammaglobin antigen (MA)
 CC in a test sample, comprising a container containing an antibody as in
 CC (A); (2) a method for producing antibodies which specifically bind to a
 CC MA, comprising administering an isolated immunogenic polypeptide or
 CC fragment to elicit an immune response, where the immunogenic polypeptide
 CC comprises at least 1 ME and has at least 50% identity to a sequence
 CC (see AAW48432) and fragments; and (3) a method for producing antibodies
 CC which specifically bind to a MA comprising administering to a mammal a
 CC plasmid comprising a sequence which encodes at least 1 ME derived from a
 CC polypeptide having an amino acid sequence (see AAW48432) and fragments.
 CC The products and methods can be used for detecting, diagnosing, staging,
 CC monitoring, prognosticating, preventing or treating, or determining
 CC predisposition to diseases or conditions of the breast such as breast
 CC cancer.

XX Sequence 93 AA;

Query Match 49.8%; Score 281; DB 19; Length 93;

Best Local Similarity 58.7%; Pred. No. 5.7e-23;

Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

OY 14 MKLIMVLMALALLHCYADSGCKLEDMVEKTI NSDISPEYKELQEFIDSDAAAEAMG 73

Db 1 mklimvmlaalsqhcysgcpllenviskrlmpgskvskellqefididnatnaid 60

OY 74 KFKOCFLNOSHRTLUKNFGMLMHTVYDSIWCNM 105

Db :|||||: ||| :| :||| :|:
 61 elkeclnqldelnsvefmqllydsslcdl 92

RESULT 13

AAV01718

ID AAV01718 standard; Protein; 93 AA.

AC AAV01718;

DT 25-JUN-1999 (first entry)

XX Mammaglobin, a mammary specific protein.

XX Human; mammary-specific protein; mammaglobin; antigen; vaccine;

XX mammaglobin-expressing cancer; breast cancer;

XX autologous tumor lymphocyte; diagnosis; marker.

XX Homo sapiens.

XX WO914230-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US17991.

XX 18-SEP-1997; 97US-0933149.

XX (UNIW) UNIV WASHINGTON.

XX Fleming TP, Watson MA;

XX WPI; 1999-244021/20.

XX N-PSDB; AAX26966.

XX Mammaglobin, secreted protein overexpressed in breast cancer

XX Claim 15; Fig 2; 60pp; English.

XX The present sequence represents a human mammary-specific protein,
 CC designated mammaglobin. The specification describes a protein
 CC comprising a mammaglobin antigen that is recognized by B and/or
 CC Tc cells specific for the natural, secreted and glycosylated form
 CC of mammaglobin polypeptide. This protein, or recombinant vectors
 CC that express it, are used in vaccines for treating mammaglobin-
 CC expressing cancers, specifically of the breast. Such cancers can
 CC also be treated using autologous tumor lymphocytes activated
 CC ex vivo with an mammaglobin antigen, then returned to the
 CC patient. Expression of mammaglobin is elevated in 27% of stage I
 CC primary breast cancers, so it represents a marker useful for
 CC diagnosis of this disease.

XX Sequence 93 AA;

Query Match 49.8%; Score 281; DB 20; Length 93;

Best Local Similarity 58.7%; Pred. No. 5.7e-23;

Matches 54; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

OY 14 MKLIMVLMALALLHCYADSGCKLEDMVEKTI NSDISPEYKELQEFIDSDAAAEAMG 73

Db 1 mklimvmlaalsqhcysgcpllenviskrlmpgskvskellqefididnatnaid 60

OY 74 KFKOCFLNOSHRTLUKNFGMLMHTVYDSIWCNM 105

Db 61 elkeclnqldelnsvefmqllydsslcdl 92

RESULT 14

AAB13786

ID AAB13786 standard; Protein; 93 AA.

XX AAB13786;

RESULT 1
US-08-821-451A-6
Sequence 6, Application US/08821451A

GENERAL INFORMATION:

APPLICANT: Jian NI, Guo-Liang Yu and Reiner Gentz
TITLE OF INVENTION: Human Endometrial Specific Steroid-
Binding Factor I, II and III
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,451A
FILING DATE: March 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,724
FILING DATE: March 21, 1996
ATTORNEY/AGENT INFORMATION:

NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-521 (PP257)
TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 95 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-821-451A-6

Query Match 87.9%; Score 496; DB 3; Length 95;
Best Local Similarity 100.0%; Pred. No. 8,8e-54;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu Jun 20 11:26:58 2002

us-09-673-39

QY 14 MKLIVTALALALHCHYADSGCKLLEDVYKTIINSISIPYKELLOEFIDSDAAEANG 73
DB 1 MKLIVTALALALHCHYADSGCKLLEDVYKTIINSISIPYKELLOEFIDSDAAEANG 73
QY 74 KFKOCFLNOSHRTLNKFGIMAHVYDSTICNKN 108
DB 61 KFKOCFLNOSHRTLNKFGIMAHVYDSTICNKN 95

PS Claim 23; Page 314; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),
CC that are highly expressed in uterine tumour tissue and which have
CC anticancer and cytostatic activity. (A) are used (i) for recombinant
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
CC are used (i) to identify agents suitable for treatment of uterine or
CC endometrial cancer; (ii) directly for treating these forms of cancer
CC (including expression from gene therapy vectors) and (iii) for
CC generation of specific antibodies. (A) are identified by assembling ESTs
CC (expressed sequence tags) from a particular tissue type before comparison
CC of expression patterns. This allows a significantly longer fragment of
CC the gene to be revealed, so should reduce the number of failures
CC associated with the fact that ESTs from different libraries may represent
CC different parts of the same unknown gene, distorting the estimated
CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent
CC protein fragments encoded by the human endometrium tumour cDNA library
CC derived EST fragments represented in AAY41981-242121.

SO Sequence 108 AA;

Query Match 100.0%; Score 564; DB 20; Length 108;

Best Local Similarity 100.0%; Pred. No. 1.5e-53;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CHARLWTDSSRLAMKLLMVLMLAALLHCYADSGCKLLEDVMEKTI NSDTSIPYKELIQ 60

DB 1 charltdssrlamkllmvlmlaallhcyadsgcklledvmeaktinsdtsipeykellq 60

QY 61 EFIDSDAAAEAMGKRFCKQFLNOSHRTIKNGLMMHTVYDSINCMKSN 108

DB 61 efidsdaaaeamgkfkqcfnglnshrtiknglmmhtvydsiwcnmksn 108

RESULT 2

ID AAM35804 standard; Protein; 95 AA.

AC AAM35804;

DT 27-MAR-1998 (first entry)

XX Human endometrial specific steroid-binding factor III.

KW Endometrial specific steroid-binding factor III; ESF III; human;

KW Clara cell secretory protein; endometrium;

KW phospholipase A2 inhibitor; polychlorinated biphenyl; antiagregant;

KW inflammation; asthma; rhinitis; cystic fibrosis; airway disease;

neoplasia; atopy; therapy; diagnosis.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Peptide 1..21

FT Protein /label= Sig_peptide

FT /label= Mat_protein

PN WO9734997-A1.

PD 25-SEP-1997.

PF 21-MAR-1996; 96WO-US03857.

PR 21-MAR-1996; 96WO-US03857.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Gentz RL, Ni J, Yu G;

WPI: 1997-480206/44.

NCSR: AAT94832.

XX Human endometrial specific steroid-binding factor I, II and III -

PT used to treat inflammation, asthma, rhinitis, cystic fibrosis,

PS Claim 19; Page 65; 92pp; English.

XX This sequence comprises human endometrial specific steroid binding
CC factor III (ESF III), a protein that inhibits phospholipase A2
CC activity, binds to polychlorinated biphenyl compounds, reduces
CC foreign protein antigenicity, inhibits monocyte and neutrophil
CC chemotaxis and phagocytosis, inhibits platelet aggregation,
CC regulates eicosanoid levels in the human uterus and controls the
CC growth of endometrial cells. The amino acid sequence was deduced
CC from a cDNA clone (see AAT94832) derived from a human endometrial
CC tumour. ESF I (see AAM35802) and ESF II (see AAM35803) are also
CC claimed. Human ESF III has about 36% identity with rat prostatic
CC steroid-binding protein. Recombinant ESF I, II and III can be
CC expressed in host cells for use in claimed methods (a) for treating
CC a patient in need of ESF I, II or III (including expression of the
CC polypeptide in vivo) and (b) for identifying compounds which bind
CC to and inhibit activation of the ESF polypeptide. ESF I, II and
CC III may be used to treat inflammation, asthma, rhinitis, cystic
CC fibrosis, airway disease, neoplasia and atopy.

SO Sequence 95 AA;

Query Match 87.9%; Score 496; DB 18; Length 95;

Best Local Similarity 100.0%; Pred. No. 2.9e-46;

Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MKLLMVLMLAALLHCYADSGCKLLEDVMEKTI NSDTSIPYKELIQEFIDSDAAAEAMG 73

DB 1 mklmvlmlaallhcyadsgcklledvmeaktinsdtsipeykellqefidsdaaaeamg 60

QY 74 KFKCFLNOSHRTIKNGLMMHTVYDSINCMKSN 108

DB 61 kfkcfnglnshrtiknglmmhtvydsiwcnmksn 95

RESULT 3

ID AAY02590 standard; Protein; 95 AA.

AC AAY02590;

DT 26-JUL-1999 (first entry)

XX A human mamoglobin homologue (HMH).

KW Human mamoglobin homologue; HMH; antagonist; neoplastic disorder;

KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;

KW teratocarcinoma; endometriositis.

OS Homo sapiens.

PN WO919487-A1.

PD 22-APR-1999.

PF 14-OCT-1998; 98WO-US21729.

PR 16-OCT-1997; 97US-0951750.

PA (INCY-) INCYTE PHARM INC.

PI Hillman JL, Murry LE, Shah P;

DR WPI: 1999-302531/25.

DR N-PSDB; AAX36138.

XX New human mamoglobin homologue (HMH), useful for diagnosing, treating